



ADAPTATION OF DECISION CRITERIA

IN VIGILANCE TASKS

James Leary

B.Sc. (Dalhousie)

M.Sc.(M.U.N.) (Halifax, Nova Scotia)

Thesis submitted to the University of Adelaide
in fulfilment of the conditions for the
Degree of Doctor of Philosophy

Department of Psychology
The University of Adelaide
South Australia

June 1983

TABLE OF CONTENTS

	Page
LIST OF FIGURES	(ix)
LIST OF TABLES	(xvii)
SUMMARY	(xx)
DECLARATION	(xxii)
ACKNOWLEDGEMENTS	(xxiii)
CHAPTER 1	
A. INTRODUCTION	1
B. REVIEW OF STUDIES ON VIGILANCE	2
1. Approaches Derived from Learning Theories	2
2. Expectancy Theory	4
3. Arousal Models	6
4. Approaches Based on Signal Detection Theory	10
(a) Assumptions of the theory	11
(b) Experimental findings related to signal detection measures	13
(c) Interpretation of β increment by the 'ideal observer' hypothesis	17
(d) Explanation of change in response latency and confidence during a vigil: the 'distance-from-criterion' hypothesis	20
(e) Problems for the ideal observer hypothesis	23
C. SOME GENERAL CONCLUSIONS	26
1. Arousal Versus a Response Optimization Interpretation of β Increment	27
2. Change in Signal Probability from Practice to Test Conditions - Problems for a Response Optimization Model	30

	Page
CHAPTER 2	32
A. EXPERIMENT ONE: EFFECTS OF A STEADY DECREASE IN SIGNAL PROBABILITY	33
1. Method	34
(a) Task and stimuli	34
(b) Apparatus	36
(c) Design	36
(d) Procedure	37
(e) Observers	39
2. Results	39
(a) Response probability	39
(b) Individual differences in the revision of response ratios	45
(c) Signal detection analysis	47
(d) Response latencies	49
3. Discussion	51
(a) Adaptive response to a change in signal probability - comparisons with previous findings	51
(b) Individual differences in the revision of response ratios	54
(c) The 'target' of revision of response ratios - general conservatism viewed as purposeful probability matching	54
(d) Individual differences in the processing of probabilistic information - the need of a further study	58
B. EXPERIMENT TWO: CONSISTENCY OF INDIVIDUAL DIFFERENCES IN THE ADAPTATION TO A DECREASING SIGNAL PROBABILITY	60
1. Method	60
(a) Observers	60
(b) Task design and apparatus	60
(c) Procedure	62

	Page
2. Results	62
(a) <u>Response Probability</u>	62
(b) Individual differences in the revision of response ratios	66
(c) Signal detection analysis	68
(d) Response latencies	69
3. Discussion	72
(a) Consistency in the degree of revision of response ratios	72
(b) Response to a change in signal probability - a proposed target for the adaptive process	73
(c) Response maintenance - a clearly inadequate interpretation of the present results	75
CHAPTER 3	81
A.1. FURTHER EVIDENCE OF NON-OPTIMAL REVISION IN β	81
2. AN ALTERNATIVE ADAPTIVE RESPONSE STRATEGY - MINIMIZE THE DISCREPANCY BETWEEN LOCAL AND CUMULATIVELY-BASED SIGNAL PROBABILITY (RESPONSE STABILIZATION)	85
3. EXPLANATION OF THE NON-OPTIMALITY OF β IN OTHER STUDIES	92
B. EXPERIMENT THREE: PERFORMANCE CHANGE FOLLOWING A PERIOD OF TIME-VARYING SIGNAL PROBABILITY	94
1. Method	97
(a) Task and stimuli	97
(b) Apparatus	98
(c) Design	99
(d) Procedure	101
(e) Observers	101

	Page
2. Results	101
(a) Response probability	102
(b) Response latencies	107
(c) Response confidence	110
(d) Signal detection measures and theoretical decision parameters	113
3. Discussion	120
(a) Relative support for the three models	120
(b) Relevance of present findings to vigilance	124
C. EXPERIMENT FOUR: EVALUATION OF THE RESPONSE STABILIZATION MODEL IN A REPRESENTATIVE VIGILANCE TASK	129
1. Method	130
(a) Task and stimuli	130
(b) Apparatus	131
(c) Design	132
(d) Procedure	132
(e) Observers	133
2. Results	133
(a) Response probability	134
(b) Response latencies	137
(c) Response confidence	141
(d) Signal detection measures and theoretical decision parameters	143
3. Discussion	146
(a) Correspondence to previous results	146
(b) The control of ongoing performance in vigilance	151

	Page
CHAPTER 4	153
A. ADJUSTMENT OF THE DECISION CRITERION THROUGH FEEDBACK	153
1. Indlin's (1976) Response Regulation Model	153
(a) Main features of the model	153
(b) Response regulation as probability-matching	155
2. Response Regulation Versus a Response Stabilization Process	156
(a) Response to change in signal probability	156
(b) A possible hierarchy of feedback systems	158
B. EXPERIMENT FIVE: DECISION CRITERION ADJUSTMENTS DURING A PROGRESSIVE DECLINE IN SIGNAL DETECTABILITY	159
1. Predictions of Three Models	160
(a) Fixed sensory cutoff model	161
(b) Response regulation	164
(c) Response stabilization	167
2. Method	172
(a) Task and stimuli	172
(b) Apparatus	173
(c) Design	174
(d) Procedure	175
(e) Observers	177
3. Results	177
(a) Response probability	177
(b) Response latencies	183
(c) Response confidence	186
(d) Differences between instruction groups	190
(e) Signal detection measures and theoretical decision parameters	194

	Page
4. Discussion	201
(a) Relative support for the three models	201
(b) Relationships among performance measures	203
(c) Relevance to previous studies in which detectability varied	207
CHAPTER 5	209
A. A CONFIDENCE-BASED FEEDBACK MODEL	209
1. General Properties of the Model	209
2. Determination of Mean Observation Distances	211
3. Response to Variation in Signal Probability	213
(a) Ramp change in signal probability	213
(b) Step change in signal probability	217
4. Vigilance Decrement	223
5. Response to Variation in Signal Detectability	225
6. A Limitation of Response Bias as a Mediator in Feedback Control	230
B. EXPERIMENT SIX: PILOT INVESTIGATION OF THE EFFECTS OF A RAMP CHANGE IN DISCRIMINABILITY	234
1. Method	234
(a) Task and stimuli	234
(b) Apparatus	235
(c) Design	236
(d) Procedure	237
(e) Observers	238
2. Results	238
(a) Pre-test differences between experimental groups	238
(b) Effects of variation in discriminability on overall performance	241
(c) Performance at the constant stimulus difference	248

	Page
3. Discussion	254
(a) Sensitivity of overall performance to the variation in discriminability	254
(b) Evidence of adaptation to variation in discriminability	255
(c) The possibility that <i>caution</i> had been adjusted	256
CHAPTER 6	262
A. CRITERION REGULATION AND GENERAL CAUTION IN RESPONDING	262
1. The Adaptive Accumulator Model	262
(a) The primary decision process	263
(b) The secondary control process	266
(c) Influence of stimulus discriminability and probability on decision criteria	269
2. Previous Results: Sensory Adaptation Versus Higher-Order Control	273
(a) Intervention by other control processes	273
(b) Change in point of sensory indifference	274
(c) Sensory bias in a data accumulation model of signal detection	280
3. A Dual-Criterion Self-Regulating Signal Detection Model	283
(a) The decision process	283
(b) Response probability, latency and confidence	285
(c) Regulation of response bias and caution	289
B. EXPERIMENT SEVEN: CRITERION ADJUSTMENTS FOLLOWING A STEP CHANGE IN STIMULUS DISCRIMINABILITY	294
1. Method	295
(a) Task and stimuli	295
(b) Apparatus	297
(c) Design	297

	Page
(d) Procedure	299
(e) Observers	299
2. Results	300
(a) Performance across all line length differences	300
(b) Performance at line length difference 3	310
(c) Early trends in performance	318
3. Discussion	320
(a) The effects of variation in caution - an anomaly	320
(b) A parallel-control model	323
(c) Previous results and the principle of 'least variation'	332
(d) Vigilance performance as affected by ISI location	340
(e) Vigilance performance as affected by ISI width	344
C. SOME GENERAL CONCLUSIONS	348
1. The Nature of Vigilance Decrement	348
2. The Proposed Model of Performance	349
REFERENCES	351

LIST OF FIGURES

		Page
Figure 1	Hypothetical distributions of the sensory effect produced by signal, s , and nonsignal, n , stimuli	11
Figure 2	Illustration of the distance-from-criterion hypothesis as applied to both response latency and confidence.	21
Figure 3	Results of Williges (1969).	24
Figure 4	Welford's (1968) model of the effect of a change in arousal on the mean and variance of the s and n distributions.	28
Figure 5	Response and stimulus presentation apparatus as used, with only minor modifications, throughout Experiments 1 to 7.	35
Figure 6	Changes in the mean conditional probability of a correct (S/s) and incorrect (S/n) signal response over the six trial blocks.	40
Figure 7	Changes in the mean unconditional probability of making a signal response compared with the local and cumulatively-based probability of a signal stimulus over the six trial blocks.	43
Figure 8	Examples of unconditional signal response probability plotted as a function of cumulatively-based signal probability.	46
Figure 9	Changes in group values of d' and β over the six trial blocks.	47
Figure 10	Signal, s , and nonsignal, n , distributions with unequal variances.	50.
Figure 11	Changes in the mean conditional probability of a correct (S/s) and incorrect (S/n) signal response over the six trial blocks.	63
Figure 12	Changes in the mean unconditional probability of making a signal response in Experiments 1 and 2 compared with the local and cumulatively-based probability of a signal stimulus over the six trial blocks.	64

	Page
Figure 13 Changes in group values of d' and β over the six trial blocks.	68
Figure 14 Changes over the six trial blocks in the latencies of correct signal (S/s), correct nonsignal (N/n), incorrect signal (S/n) and incorrect nonsignal (N/s) responses.	70
Figure 15 Effect of a change from an intermediate to a more lax decision criterion.	79
Figure 16 Flow diagram of the proposed response stabilization process for the minimization of discrepancies between the local (L) and cumulatively-based (C) probability of making a signal response.	87
Figure 17 The general form of changes in the objective local (L) and cumulative (C) probability of a signal from training to testing.	88
Figure 18 Changes in local (L) and cumulatively-based (C) probability of signal occurrence, and the difference ($L-C$), when signal probability is reduced in a ramp fashion and then held constant.	95
Figure 19 Changes in β predicted by each of three theoretical models for the two situations of time-varying signal probability depicted in figure 18.	96
Figure 20 The probability that the longer line would occur to the right (signal probability) during training and each of the twenty successive trial blocks.	100
Figure 21 Changes in the mean conditional probability of (a) correct and (b) incorrect signal responses over the five temporal periods within each experimental condition.	104
Figure 22 Changes in the mean unconditional probability of making a signal response over the five temporal periods within each experimental condition.	105
Figure 23 Changes in the mean unconditional probability of making a correct response (signal or nonsignal) over the five temporal periods within each experimental condition.	106

	Page
Figure 24 Changes in the mean latencies of (a) correct signal and (b) correct nonsignal responses over the five temporal periods within each experimental condition.	108
Figure 25 Changes in the mean confidence expressed in (a) correct signal and (b) correct nonsignal responses over the five temporal periods within each experimental condition.	111
Figure 26 Changes in group values of β over successive blocks of 100 trials during the experimental session.	115
Figure 27 A comparison of changes in group values of β and mean values of $(1-c)$ during the experimental session in each condition.	116
Figure 28 A comparison of changes during the session in group values of β and mean values of $(1-c)$.	122
Figure 29 Hypothetical distributions of the sensory effect produced by signal and nonsignal stimuli.	125
Figure 30 Examples of (a) a nonsignal and (b) a signal stimulus as used in Experiments 4 and 5.	130
Figure 31 Changes in the mean conditional probability of (a) correct and (b) incorrect signal responses over the twenty trial blocks.	135
Figure 32 Changes in the mean unconditional probability of making (a) a correct response and (b) a signal response over the twenty trial blocks.	136
Figure 33 Changes in the mean latencies of (a) correct signal and (b) correct nonsignal responses over the twenty trial blocks.	139
Figure 34 Changes in the mean confidence expressed in (a) correct signal and (b) correct nonsignal responses over the twenty trial blocks.	142
Figure 35 Changes in the observed group values of β compared with the optimal values specified by signal detection theory over the twenty blocks.	144

	Page
Figure 36 A comparison of changes in group values of β and mean values of $(1-c)$ over the twenty trial blocks.	145
Figure 37 Diagram of a simple feedback system in which the present output (y) of the system is compared with the input (x) which has been applied to it.	154
Figure 38 The fixed sensory cutoff model.	162
Figure 39 Patterns of change in performance that are expected to follow from a downward movement of the signal distribution past a fixed sensory cutoff, as shown in figure 38.	163
Figure 40 The response regulation model.	165
Figure 41 Patterns of change in performance that are expected to follow from the response regulation process shown in figure 40.	166
Figure 42 The adjustment in the observer's decision criterion expected by the response stabilization model to follow from an initial tendency to maintain a fixed sensory cutoff.	169
Figure 43 Patterns of change in performance that are expected to follow from an initial tendency to maintain a fixed sensory cutoff followed by a tendency toward response stabilization, as shown in figure 42.	170
Figure 44 The level of detectability at which signals were presented during the four blocks of training trials and the twenty successive blocks of experimental trials in Experiment 5.	175
Figure 45 Changes in the mean conditional probability of (a) correct and (b) incorrect signal responses over the twenty trial blocks for each group of observers.	179
Figure 46 Changes in the mean unconditional probability of making (a) a signal and (b) a correct (signal or nonsignal) response over the twenty trial blocks for each group of observers.	181

	Page
Figure 47 Changes in the mean latencies of (a) correct signal and (b) correct nonsignal responses over the twenty trial blocks for each group of observers.	185
Figure 48 Changes in the mean confidence expressed in (a) correct signal and (b) correct nonsignal responses over the twenty trial blocks for each group of observers.	188
Figure 49 Changes in group values of d' and β over the twenty trial blocks for each group of observers.	195
Figure 50 A comparison of changes over the twenty trial blocks in the value of β .	197
Figure 51 Values of the decision criterion index suggested by Indlin (1976) across the twenty trial blocks.	200
Figure 52 The relationship between response latency, given by the dashed line, and the distance of an observation from the decision criterion if there is an intensity component of response latency that is inversely proportional to sensory activity.	206
Figure 53 The average distance at which observations lie from the decision criteria as signal probability declines from .5 when d' is constant at 2.0 and the criterion remains fixed at the intersection of the s and n distributions.	214
Figure 54 The mean observation distances of signal (S) and nonsignal (N) responses for each trial block of Experiment One and test period of Experiment Four.	216
Figure 55 The rate at which the output of a second-order feedback control system approaches its target level (T) under various values of the damping ratio.	220
Figure 56 Mean observation distances of signal (S) and nonsignal (N) responses for values of d' between 1.0 and 4.0.	226
Figure 57 The mean observation distances of signal (S) and nonsignal (N) responses at each level of detectability in Experiment Five.	228

Figure 58	Hypothetical distributions of noise, n , a high intensity signal, s_1 , and a low intensity signal, s_2 , in the experiment of Gescheider, Wright and Polak (1971).	229
Figure 59	Hypothetical distributions of differences in sensory effect arising in a discrimination task in which either the left, L , or right, R , of two sources is of the greater magnitude.	232
Figure 60	The basic design of Experiment 6.	237
Figure 61	The mean conditional probability of making (a) a correct right and (b) a correct left response across the line length differences at each level of discriminability for each group of observers.	242
Figure 62	The mean latencies of (a) correct right and (b) correct left responses across the line length differences at each level of discriminability for each group of observers.	245
Figure 63	The mean confidence expressed in (a) correct right and (b) correct left responses across the line length differences at each level of discriminability for each group of observers.	247
Figure 64	The mean conditional probability of making (a) a correct right and (b) a correct left response to stimuli with a line length difference of 5 at each level of discriminability for each group of observers.	249
Figure 65	The mean latencies of (a) correct right and (b) correct left responses to stimuli with a line length difference of 5 at each level of discriminability for each group of observers.	251
Figure 66	The mean confidence expressed in (a) correct right and (b) correct left responses to stimuli with a line length difference of 5 at each level of discriminability for each group of observers.	253
Figure 67	Hypothetical critical totals, $C+$ and $C-$, required in order to reach a respective right or left decision.	258

Figure 68	Flow diagram of the primary decision process in Vickers' adaptive accumulator model.	264
Figure 69	A branch of the secondary control process in Vickers' adaptive accumulator model.	268
Figure 70	Hypothetical distributions of sensory differences produced by a stimulus with a greater magnitude to the right, distributed as r , and a stimulus with a greater magnitude to the left, distributed as l .	276
Figure 71	Hypothetical distributions of sensory differences produced by a stimulus with a greater magnitude to the left.	279
Figure 72	Hypothetical distributions of the sensory effect produced by signal, s , and nonsignal, n , stimuli in the 'interval of uncertainty' model.	284
Figure 73	The distance-from-criterion hypothesis applied to the 'interval of uncertainty' (IU) detection model.	287
Figure 74	Hypothetical distributions of signal, s , and nonsignal, n , stimuli.	289
Figure 75	Hypothetical distributions of s and n when there is a decline in the magnitude of s , so that it becomes distributed as s' , and an increase in the magnitude of n , so that it becomes distributed as n' .	293
Figure 76	An example of a variable (left) and standard (right) display as was presented to the observer on each trial of Experiment 7.	296
Figure 77	The line length differences which were presented during each trial block of Experiment 7.	298
Figure 78	The mean probability of making a correct response across the line length differences in each block of 48 trials.	301
Figure 79	The mean latency of all responses across the line length differences in each block of 48 trials.	304

Figure 80	The mean confidence expressed in correct responses across the line length differences in each block of 48 trials.	307
Figure 81	The mean confidence expressed in incorrect responses across the line length differences in each block of 48 trials.	309
Figure 82	The mean probability of making a correct response to stimuli with a line length of 3 in each block of 48 trials.	311
Figure 83	The mean latency of all responses to stimuli with a line length difference of 3 in each block of 48 trials.	313
Figure 84	The mean confidence expressed in all responses to stimuli with a line length difference of 3 in each block of 48 trials.	315
Figure 85	A dual-criterion detection model in which confidence, given by the dashed line, is assumed to be a function solely of sensory magnitude.	322
Figure 86	Decision and sampling processes in the parallel-control model.	324
Figure 87	Relationship between the width of the interval of sensory indifference, ISI.	326
Figure 88	Relationship between the width of the interval of sensory indifference, ISI.	331
Figure 89	Hypothetical distributions of the sensory effect produced by signals, s , and nonsignals, n .	334
Figure 90	Hypothetical distributions of the sensory effect produced by signals, s , and nonsignals, n .	337
Figure 91	The initial position of an ISI.	341

LIST OF TABLES

		Page
Table 1	Results of analyses of variance on conditional response probabilities.	41
Table 2	Results of analyses of variance on signal detection measures.	48
Table 3	Mean latencies for the four categories of response.	49
Table 4	Mean latencies for the four categories of response.	72
Table 5	Results of analyses of variance on response probabilities.	103
Table 6	Results of analyses of variance on response latencies.	108
Table 7	Mean latencies and confidence for the four categories of response.	109
Table 8	Results of analyses of variance on response confidence.	111
Table 9	Results of analyses of variance on signal detection measures.	114
Table 10	Pearson product-moment correlations between the mean value of $(1-c)_i$ in each block of 100 test trials and the group value of β , either in that trial block (β_i) or the following trial block (β_{i+1}).	118
Table 11	Results of analyses of variance on response probabilities.	134
Table 12	Results of analyses of variance on response latencies and confidence.	138
Table 13	Mean latencies and confidence for the four categories of response.	140
Table 14	Results of analyses of variance on signal detection measures.	143
Table 15	Line lengths in the central matrix of signal and nonsignal displays.	173
Table 16	Results of analyses of variance on response probabilities over all levels of detectability.	178

	Page
Table 17 Results of analyses of variance on response probabilities over detectability levels 5 to 7.	180
Table 18 Results of analyses of variance on response latencies.	184
Table 19 Mean latencies and confidence for the four categories of response among all observers over detectability levels 1 to 4.	187
Table 20 Results of analyses of variance on response confidence.	187
Table 21 The number of trial blocks in which each instruction group had the highest mean value of each of the performance measures shown in Figures 45 to 48.	191
Table 22 Mean measures of performance for the two instruction groups during the final block of training trials.	193
Table 23 Results of analyses of variance on signal detection measures over detectability levels 1 to 4.	194
Table 24 Results of analyses of variance on the data from the second block of practice trials.	239
Table 25 Results of analyses of variance on response probabilities averaged across the line length differences at each level of discriminability.	241
Table 26 Results of analyses of variance on response latencies averaged across all line length differences at each level of discriminability.	244
Table 27 Results of analyses of variance on response confidence averaged across all line length differences at each level of discriminability.	246
Table 28 Results of analyses of variance on response probabilities at line length difference 5.	248
Table 29 Results of analyses of variance on response latencies at line length difference 5.	250

Table 30	Results of analyses of variance on response confidence at line length difference 5.	252
Table 31	Expected performance measures when the means of the <i>s</i> and <i>n</i> distributions are separated by 2.0 standard deviation units and an unbiased IU is expanded from a width of 0.0 (single criterion) to 3.0 standard deviation units.	291
Table 32	Results of analyses of variance on correct response probability averaged across the three line length differences in each block of trials.	303
Table 33	Results of analyses of variance on response latencies averaged across the three line length differences in each block of trials.	305
Table 34	Results of analyses of variance on response confidence averaged across the three line length differences in each block of trials.	308
Table 35	Results of analyses of variance on the probability of making a correct report of a stimulus with a line length difference of 3.	312
Table 36	Results of analyses of variance on the latencies of responses to stimuli with a line length difference of 3.	314
Table 37	Results of analyses of variance on the confidence of responses to stimuli with a line length difference of 3 over trial blocks 7 to 14.	316
Table 38	Mean performance measures during trial blocks 1 and 2 (median discriminability)	318

SUMMARY

Adaptation of decision criteria in vigilance tasks

Decrements in performance during vigilance tasks are usually associated with an increased strictness in the criterion for a signal response. In order to determine whether this can be attributed to a drop in signal probability between training and testing, in Experiments 1 and 2 signal probability was reduced in a ramp fashion during the trial sequence of a discrimination task. Contrary to the predictions of signal detection theory, bias increased toward making signal responses in a manner which also disconfirmed an alternative hypothesis that observers will maintain an initial proportion of signal responses. Instead, observers appeared to adjust their decision criteria so as to maintain the local probability of a signal response at the level of cumulatively-based signal probability. This response stabilization model received further support in Experiment 3 in which signal probability was held constant following a ramp decline during a discrimination task. These results were replicated in a task more appropriate to vigilance in Experiment 4.

In Experiment 5 signal detectability was varied, while *a priori* probabilities were held constant. Observers did not maintain a fixed sensory cutoff, in order to satisfy the Neyman-Pearson objective, nor vary the cutoff according to a continual regulation process, in order to match

response and stimulus probabilities. Instead they appeared initially to maintain a fixed cutoff, which varied afterward, according to a stabilization process. A confidence-based model was introduced, incorporating response stabilization, and according to which observers adjust their decision criteria so as to maintain observations at some constant average distance from the decision criterion. As in the previous single-criterion detection models, control is mediated through variation in response bias, with the result that no compensation would be predicted when symmetrical effects occur for the two response alternatives. In Experiment 6, such effects were produced by a ramp decline in stimulus discriminability. Changes in latency to a constant amount of stimulus difference indicated that there had been adaptation consistent with a change in the general level of response caution. In Experiment 7 stimulus discriminability was varied in a recurring step fashion. In agreement with the adaptive accumulator model of Vickers (1979), caution was inversely related to the level of discriminability. While a reformulation of a sequential decision model of signal detection can also account for this relationship, neither model can explain the direct relationship between speed and accuracy which followed a step change in discriminability. A model is presented in which sensory boundaries are adjusted to maintain the rate at which information enters the decision process, while control accumulators ensure that decisions are made at target levels of confidence. The model can account for a number of seemingly contradictory trends found to occur over time in vigilance tasks.

DECLARATION

This thesis contains no material which has been accepted for the award of any other Degree or Diploma in any University, and to the best of my knowledge and belief, it contains no material previously published or written by any person, except when due reference is made in the text.

Signed

James Leary

June, 1983

ACKNOWLEDGEMENTS

I wish to express my sincere, and deep, gratitude to Doug Vickers, not only for carefully, and expertly, supervising all the research presented in this thesis, but also for his role as advisor and personal confidant. His encouragement and advice will never be forgotten. Thanks are also due to Peter Barnes, who, with impressive competence and often at very short notice, carried out the computer programming required to control all aspects of the experiments, and to perform the preliminary data analyses. Lastly, I wish to express my appreciation to Margaret Blaber, for her careful typing of this manuscript, and to Judy Fallon, who skillfully drew up the figures and tables.

The execution of the computing and experimental work for this thesis was materially assisted by a grant to D. Vickers from the Australian Research Grants Committee.



CHAPTER 1

A. INTRODUCTION

Opinions regarding human vigilance seem to have polarized into two main theoretical camps and, unfortunately, there has been no concerted effort towards integration. On the one hand, some theorists believe that the ability to maintain a high level of vigilance performance depends solely on overall lack of sensory excitation, arousal, or some other physiological variable. This seems to regard the human being as a passive sensing device. On the other hand, some theorists, while acknowledging the role of the overall efficiency level of the observer, tend to explain all aspects of vigilance behaviour in terms of decision-making variables. Past experience with the task situation, and knowledge of the crucial signals to be presented in it, thus play a more important role than organismic considerations.

Despite the origins of this thesis in the field of vigilance, the outcome of the experiments reported in the next chapter led the writer to consider, and examine, topics which were somewhat removed from those previously considered within the confines of vigilance research. However, the model which is ultimately presented in the final chapter not only has the capacity to be applied to vigilance, but also specifies independent sensory and decision processes which, through their individual or joint action, can potentially account for a number of empirical vigilance phenomena. It is therefore hoped that a useful synthesis of views has been achieved. In order to trace the development of the research to be presented later, a brief review of some aspects of vigilance performance will be given in the present chapter. This review is not meant to be exhaustive since excellent reviews already exist in J. Mackworth (1969, 1970), Davies and Tune (1970), Broadbent (1971), and, more recently, in the proceedings of a N.A.T.O. symposium on the topic (Mackie, 1977).

B. REVIEW OF STUDIES ON VIGILANCE

Research on vigilance originally developed during the Second World War as the result of the obvious practical need to understand and predict the performance of military radar observers or sonar operators at detecting the faint improbable signals which indicated the presence of enemy aircraft or ships. In the first published study on the topic, N. Mackworth (1948) defined vigilance as "a psychological readiness to perceive and respond, a process which, unlike attention, need not be consciously experienced" (p. 6). At this stage the phenomenon of greatest interest was the apparent deterioration in the ability to detect signals as time progressed in a vigil. In the three decades which followed there have been almost as many theories advanced to explain this apparent decrement as there are factors appearing to affect it. We shall first look at the 'traditional' concepts which have been used to explain vigilance decrement, and then examine in more detail the relatively recent applications of signal detection theory.

1. Approaches Derived From Learning Theories

One common assumption among the early theories was that the supposed vigilance decrement reflected changes in some central process rather than in the peripheral sensory systems. This assumption was based partly on the observation that a decline in efficiency could be prevented if observers were supplied with knowledge of results (KR) about their detection performance (N. Mackworth, 1950; J. Mackworth, 1964). A similar "recovery of vigilance" was found following periods of inactivity (or rest pauses) interpolated during the session (N. Mackworth, 1950; Adams, 1956; Jenkins, 1958; Colquhoun, 1959).

Such a return of detection performance to its initial level was first explained by N. Mackworth (1950) in terms of learning theory. Specifically, Mackworth proposed that deterioration in performance was the result of the development of a state of internal inhibition. On this view the detection response was regarded as a conditioned response, while the knowledge of current detections was taken to constitute reinforcement of this response. It was then argued that without the provision of knowledge of results as a reinforcing agent, inhibition would build up and the detection response would extinguish. Since interpolated periods of inactivity would allow such effects of inhibition to dissipate, the extinguished response should be restored following such periods. Similar hypotheses, expressed in terms of the Pavlovian concept of classical conditioning and the Hullian concept of reactive inhibition, were also put forward by Broadbent (1953) and Adams (1956).

Still within a learning theory framework, a somewhat different approach was advanced by Holland (1958, 1963) who stressed the operant conditioning of orienting, or observing, responses (ORs), which were directed toward the display on which signals were presented. Usually the OR is defined as a simple overt response, such as a button press, which is required to make the stimulus display visible for a certain period of time. On Holland's approach, the OR, rather than the overt detection response, is treated as an instrumental response, which is reinforced by the detection of signals. The rate at which ORs are emitted should then parallel the frequency of signal detections, and the decrement should be accompanied by a decline in the OR rate.

Contrary to this view, however, ORs have been found to increase over time with a decrease in detection performance (Guralnick, 1972). Moreover, in two studies which determined whether the observer was attending to the

signal source, Baker (1960b) and N. Mackworth, Kaplan and Metley (1964) have shown that missed signals do not parallel the frequency with which observers look away from the display if one source is monitored. However, in a "two source" situation most of the signals missed were not fully fixated, and the probability of detection when the signal was fully fixated was the same as the single source situation where all the signals, missed as well as detected, were fixated (N. Mackworth et al., 1964). The role of orienting responses would thus appear to be a decisive factor only when more than one source of input has to be monitored.

2. Expectancy Theory

Other evidence inconsistent with the early learning theories comes from work on the role of the inter-signal interval by Adams and others. Adams (1956) described a vigilance task as a situation in which "an otherwise steady stimulus state is aperiodically punctuated by stimulus change" (p. 204). On this view one of the essential features of vigilance tasks is the temporal uncertainty in the appearance of signals. The importance of this variable in determining the efficiency of vigilance performance was emphasized in the expectancy model of Deese (1955) and its later elaboration by Baker (1959, 1963). It is assumed that previous experience with the course of events in the task leads the observer to develop an expectancy about when signals are due to occur. It is further assumed that the level of expectancy determines the degree of vigilance and hence the probability that a signal will be detected. Further, expectancy fluctuates in a rational manner so that an expectancy function is formulated that "drops to a low level after each signal, mounts rapidly as the next signal is expected, and then tends to taper off if the expectancy is not confirmed" (Baker, 1959, p. 37). The peak of this

function is assumed to occur when a period of time has elapsed that is equal to the mean interval between previous signal reports. This view opposes the notion of an inhibitory process, in that vigilance is considered to be a state of excitation determined by subjective expectancies.

Research into the effects on vigilance performance of the regularity and length of the interval between signals has led to findings consistent with the expectancy hypothesis. Baker (1959) gave evidence that performance is highest when a period of time equal to the mean inter-signal interval has elapsed since the previous detection, which then coincides with the peak of the hypothesized expectancy function. Performance is also found to improve when the mean inter-signal interval is made smaller by increasing the rate at which signals are presented (Deese, 1955; Jenkins, 1958; Kappauf & Powe, 1959; Baker, 1960a; Bowen, 1964). Presumably this is the result of more accurate expectancies, since a higher signal rate provides observers with a larger sample of intervals upon which to base estimates of the mean inter-signal interval (Baker, 1959). Also, as first suggested by Deese (1955), when the observer expects more signals, a general increase in the level of expectancy could be assumed to occur.

On an expectancy hypothesis, vigilance performance will also decline if the regularity of signals decreases due to a greater temporal randomization of signal presentations (Adams & Boulter, 1964; Smith, Warm & Alluisi, 1966; Warm & Alluisi, 1971). This follows from the model, since a decrease in signal regularity means that fewer signals will appear at the mean inter-signal interval, at which expectancy is highest. Hence, the proportion of missed signals will increase, and the number of confirmed expectancies will decrease, so that perceived signal rate will decline. This leads in turn to a greater inaccuracy in estimates of the mean inter-signal interval, as well

as a decline in the general level of expectancy. Detection performance will then progressively deteriorate in a "vicious circle" effect.

The expectancy model thus provides adequate explanations of the effects of signal rate and regularity. In contrast, in the learning models, the build up of inhibition should be cumulative with repeated detections, provided there is no KR. Increased signal rate should then lead to decreased performance efficiency since the rate of extinction should be greater. The learning models also have difficulty in explaining the decrease in detection efficiency for inter-signal intervals longer than the mean (Baker, 1959; Kappauf & Powe, 1959; Johnston, Howell & Goldstein, 1966) since, according to these models, a longer than average inter-signal interval would allow more time for inhibition to dissipate.

3. Arousal Models

While the expectancy approach appears to provide a more valid interpretation of vigilance than the earlier learning models, it in its turn encounters difficulty in explaining some findings on the effects of KR. On the expectancy hypothesis, when all signals are accompanied by KR, performance should improve due to a more accurate representation of the temporal sequence of events and subsequent confirmation of expectancies. However, beneficial effects have been found when KR is given about a secondary task rather than the main vigilance task (Baker, 1961). In such a situation expectancy about the main task should not be affected. Beneficial effects have also been found when less than half the signals are accompanied by KR (Johnson & Payne, 1966). Here an expectancy model would have predicted a detrimental effect, as perceived signal rate should have declined, due to too few confirmations of expectancy; that is, a falsely-low expectancy should have developed.

One possible explanation of these discrepancies lies in the postulation of some additional motivational variable, since KR appears to have a beneficial effect on performance beyond merely providing knowledge of the temporal or spatial features of a signal. Evidence for such an addition is also provided by the demonstration of an "end-spurt" effect (Bergum & Lehr, 1963; Catalano, 1973), in which vigilance performance improves toward the end of a session when observers are aware that it is about to finish. Such motivational effects can be interpreted within an arousal model of vigilance, in which performance is seen as the reflection of the overall activation level, or arousal, of the individual. For example, Deese (1955) draws an analogy between Kleitman's (1939) concept of "wakefulness of choice" and the ability to remain alert in a vigilance task. Deese's approach proposes that this ability depends "not only upon the autonomous activity of the cortex, but upon the level of sensory input as well" (1955, p. 367). The typically monotonous, unstimulating environment of the vigilance task represents a reduction in sensory input to the organism and so results in a general decline in the level of alertness, or arousal, of the observer. KR or the knowledge that a session is about to end has an arousing effect and can counteract this decline.

Corcoran (1962a) has gone on to advance the operational definition of arousal as the "inverse probability of falling asleep". This definition, combined with the inverted-U relationship that theorists claim to exist between arousal and performance (e.g. Hebb, 1955; Malmö, 1959), has met with much success in explaining the effects, and interactions, of various stressors which have been found to influence vigilance performance. In agreement with the definition, Corcoran (1964a) presented physiological evidence that sleep deprivation is a state of lowered arousal. Wilkinson (1958) has also reported that individuals deprived of sleep for 30 hours show a greater tendency toward

performance decrement during a vigilance task than do normals. It has since been demonstrated that vigilance performance is particularly sensitive to even mild degrees of sleep deprivation (Wilkinson, Edwards & Haines, 1966; Hamilton, Wilkinson & Edwards, 1972).

On the other hand, tasks having higher "interest", in the sense of motivational or arousing value (rather than merely increased complexity), are relatively insensitive to sleep deprivation (Wilkinson, 1964). In an analogous manner, Corcoran (1962b) found that 90 dB noise reduced the greater tendency toward performance decrement characteristic of sleep loss. The effect of sleep loss is also lessened when signal frequency is doubled through an increase in the rate of occurrence of all events, signal and nonsignal (Corcoran, 1963). The signal rate effect is thus seen as the result of additional incoming stimulation, which has an arousing influence and combats the effect of sleep loss. Noise could be assumed to operate in a similar way.

A number of other findings are consistent with an arousal hypothesis. For example, under conditions of normal sleep, the ingestion of stimulant drugs prevents the tendency toward vigilance decrement, while depressant drugs enhance it (N. Mackworth, 1950; Treadwell, 1960; Bakan, 1961; Loeb, Hawkes, Evans & Alluisi, 1965; Keister & McLaughlin, 1972). Again, vigilance performance is superior later in the waking day when arousal would be expected to be near its peak (Blake, 1967), and is further affected by individual differences in characteristic levels of arousal (Corcoran, 1962a, 1964b, 1965). Specifically, Extraverts, who appear to be chronically less aroused than Introverts, show a greater tendency toward vigilance decrement (Bakan, Belton & Toth, 1963). The addition of a secondary task, as an arousing influence, then benefits Extraverts more than Introverts (Bakan, 1959), as is also the case with an increase in signal frequency (Colquhoun, 1961).

Among Extraverts the greater tendency toward decrement is reduced by the addition of noise (Davies & Hockey, 1966). Lastly, Extraverts show a preference for varied auditory stimulation during a vigilance task while Introverts show a preference for its absence (Davies, Hockey & Taylor, 1969).

These findings are all consistent with the **hypothetical inverted-U** relationship between arousal and vigilance performance. According to this approach arousing influences, such as noise, would tend to cancel the effects of de-arousers, such as sleep loss, on performance. Combination of arousers would have the effect of putting performance beyond its optimal level at the peak of the inverted-U so that no change, or poorer performance, results. An example would be the relative lack of improvement in the performance of Introverts when signal frequency is increased. The model also receives support from results obtained with the response latency measure of vigilance efficiency. This measure has been considered to be a more sensitive index of vigilance level than the more frequently reported measure of detection efficiency, that is, the percentage of signals detected (Buck, 1966). A change in the features of a vigilance task may have no effect on the detection measure, but a marked effect on the latencies of detections (e.g. Warm & Alluisi, 1971; Davies, Lang & Shackleton, 1973).

Where latencies have been included in analysis the findings have usually been favourable for arousal theory. Detection latency has been found to show a general increase with time in a vigil, independent of different experimental manipulations (Jenkins, 1958; Adams & Boulter, 1962; Loeb et al., 1965; Colquhoun, 1969; Lisper, 1969; Warm & Alluisi, 1970; Lisper, Kjellberg & Melin, 1972), which indicates a constant decline in the level of alertness. Such a decline in alertness has been shown in a monitoring task of only 10 minutes' duration (Lisper & Ericsson, 1973). Detection latency

has also been found to be inversely related to signal frequency (Johnston et al., 1966; Smith et al., 1966; Martz, 1967; Johnston, Howell & Williges, 1969; Lisper, 1969; Warm & Alluisi, 1971; Lisper & Ericsson, 1973), and its increment with time is greatest at low signal frequencies (Johnston et al., 1966). Reaction time increment, like detection decrement, can be alleviated by the action of a stimulant drug (Loeb, et al., 1965). Detection latency is also inversely related to event rate in a vigilance task (Colquhoun, 1969) in agreement with the assumed importance of gross sensory bombardment on vigilance level.

4. Approaches Based on Signal Detection Theory

While findings with the detection latency measure have been consistent with an arousal interpretation, a study by Parasuraman and Davies (1976) which recorded latencies of "nonsignal" responses during a vigilance task does not support the approach. This study departed from the typical vigilance experimental design in that observers responded to all events rather than make "signal" responses only. The usual increase in signal response latency occurred as time passed in the vigil, but this was accompanied by a significant decline in the latencies of nonsignal responses.

This finding is contrary to the notion of a general decline in arousal or alertness, which would have resulted in increased latencies for all responses. The result, however, is consistent with a theory that has been applied to vigilance and has cast doubt on the existence of a time-dependent vigilance decrement, an assumption held by all of the early theories mentioned. The theory of signal detection, advanced by Tanner and Swets (1954), Swets, Tanner and Birdsall (1961) and Green and Swets (1966), has become increasingly popular in vigilance research, at least as a descriptive adjunct to other

theoretical approaches. The theory allows for separate, presumably independent, measures of the ability of the observer to discriminate between signal and nonsignal stimuli and the bias which the observer has towards making a signal report.

(a) Assumptions of the theory

The theory proposes that an observation, x_i , is made during a fixed interval of time and on this basis the observer makes a decision as to whether it arose from random background activity ("noise") or the addition of a signal

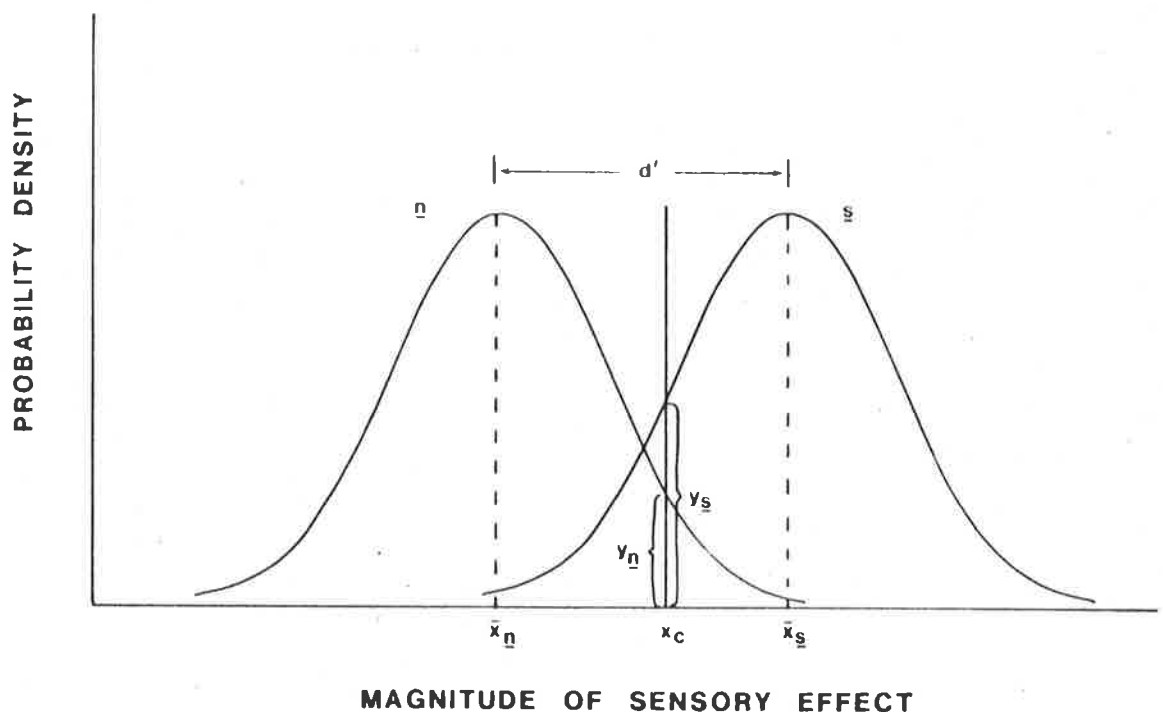


FIGURE 1. Hypothetical distributions of the sensory effect produced by signal, s , and nonsignal, n , stimuli. The distance, d' , between the means, $\bar{x}_s$ and $\bar{x}_n$, respectively, of the s and n distributions is a measure of the discriminability of the signal. Response bias is measured by the ratio of y_s to y_n , where y_s and y_n are the respective ordinate values of the s and n distributions at the cutoff position, x_c .

to that activity ("signal plus noise"). The probability that the observation could have arisen from noise alone is assumed to take the form of a Gaussian distribution, n , and that it arose from the addition of a signal that of a second Gaussian distribution, s , as shown in Figure 1. In the usual vigilance task, the noise distribution can be assumed to represent the sensory effect produced by nonsignal stimuli, where the signal is defined as an increment in intensity along some dimension above that of nonsignal stimuli. Thus the means of the signal and noise distributions are separated by a distance that is proportional to the strength of the signal stimulus, and it is usually assumed that the distributions have equal variances. It is further assumed that the observer places a cutoff at some point, x_c , along the sensory continuum such that observations which give rise to amounts of evidence which equal or exceed x_c will be reported as a signal, and those of amounts less than x_c will be reported as a nonsignal. Responses made on the basis of observations lying to the right of the cutoff can then be classified as correct detections, S/s , and false alarms, S/n , and to the left of the cutoff as correct rejections, N/n , and misses, N/s .

Provided that the two distributions are Gaussian and have equal variances, the distance between them, d' , is given by

$$d' = Z_{S/n} - Z_{S/s} \quad (1)$$

where $Z_{S/n}$ and $Z_{S/s}$ are the standard normal deviates of the n distribution and s distribution that are associated with a particular pair of false alarm and correct detection probabilities, respectively. The index, d' , is a measure of the discriminability of the signal or, alternatively, the observer's sensitivity. The ratio of the ordinate value of the s distribution to that of the n distribution at x_i gives the **likelihood that the observation arose from a signal** as opposed to a nonsignal. The value of this ratio at x_c gives a measure of the observer's criterion, β , which is inversely related to the amount of

bias the observer has toward making a signal report.

If different values and costs are attached to the four response categories the expected value of a decision will be maximised by adopting an optimal value of β , such that

$$\beta = \frac{p(n)}{p(s)} \cdot \frac{V_{Nn} + C_{Sn}}{V_{Ss} + C_{Ns}} \quad (2)$$

where $p(n)$ and $p(s)$ are the a priori probabilities of noise, or nonsignals, and signals, respectively; V_{Nn} and V_{Ss} are the values of a correct rejection and detection, respectively; and C_{Sn} and C_{Ns} are the costs of a false alarm and miss, respectively. Where the payoff matrix is symmetrical the percentage of correct responses is maximised by adopting

$$\beta = \frac{p(n)}{p(s)} \quad (3)$$

(b) Experimental findings related to signal detection measures

As d' is the measure of some underlying discriminative capacity and β is that of some judgemental criterion, **a willingness to respond, it has been** suggested by Baddeley and Colquhoun (1969) that β might be used "as a measure of the subject's expectancy" (p. 176) and so could be considered the predictive index of expectancy theory. Earlier evidence for such a view was provided by Colquhoun (1961, 1966) who showed that the signal rate effect could be largely attributed to an increase in signal probability. In his task Colquhoun (1961) presented observers with two types of event: signal events and nonsignal, or "unwanted", events. At a constant event rate detection performance was higher when the probability that an event would be a signal was higher. Further, at a constant signal rate performance was greater if

fewer events were presented, that is, when signal probability was higher. At a constant signal probability, however, detection performance was virtually independent of event rate and, therefore, of signal rate. Hence detection efficiency depended more on the probability that an event would be a signal than on the actual frequency with which signals were presented. Consistent with this interpretation the improvement in detection rate which follows an increase in signal probability is accompanied by an increase in false alarm rate. This increase is associated with a decrease in β (Broadbent & Gregory, 1965; Baddeley & Colquhoun, 1969; Williges, 1971; Murrell, 1975; Parasuraman & Davies, 1976), and would be predicted by the inverse relationship between β and signal probability that is specified by equation 3.

According to signal detection theory, a decrement in detection probability during a vigilance task can be the result of either an increase in β or a decrease in d' . In the first published paper which analyzed vigilance within a signal detection theory perspective, Egan, Greenberg and Shulman (1961) suggested that decrement was attributable to a progressive increase in the criterion for making a signal response. Consistent with this interpretation, earlier studies (e.g. Bakan, 1955; Adams, 1956) had reported decreases in false alarms during a session, but these responses were not usually incorporated into data analyses. Since the study of Egan et al. signal detection analyses of vigilance decrement have led to the prevalent finding of an increase in β , or decrease in the bias toward a signal response, as time passes in a vigil (Broadbent & Gregory, 1963, 1965; Loeb & Binford, 1964, 1968; Jerison, Pickett & Stenson, 1965; Loeb et al., 1965; Binford & Loeb, 1966; Levine, 1966; Colquhoun & Baddeley, 1967; J. Mackworth, 1968; Baddeley & Colquhoun, 1969; Colquhoun & Edwards, 1970; Guralnick & Harvey, 1970; Thurmond, Binford & Loeb, 1970; Williges, 1971, 1973; Guralnick, 1972; Milošević, 1974; Murrell, 1975; Parasuraman & Davies, 1976).

While decrements in d' during a vigil have been reported, they have usually occurred in conjunction with an increase in β and appear to depend upon a high demand being placed on the particular sensory systems involved, such as the requirement of continuous viewing in a visual task. For example, J. Mackworth and Taylor (1963) found decreases in d' accompanied the detection decrement which they observed in tasks requiring the report of a brief pause in the continuous movement of the second hand of a clock. However, no mention was made of any change which may have taken place in β . J. Mackworth (1965) presented events at rates of 40 or 200 per minute in a visual task and found a decrease in d' only at the higher event rate. Although β was not measured, declines in both detection and false alarm probabilities suggest that β did increase.

In a later study, J. Mackworth (1968) compared performance on the "continuous" clock, as used in the Mackworth and Taylor study, with the "jump" clock originally used by N. Mackworth (1950), on which there were discrete movements of the second hand at the rate of one per second. Decreases in d' occurred only with the continuous clock, whereas β showed a tendency to increase in both tasks. In a visual task with event rates which ranged from 6 to 24 per minute, Loeb and Binford (1968) found d' decrements which appeared more marked at the higher event rates, but found no such decrements in an associated auditory task. However, there were general increases in β in both types of task at all event rates. Parasuraman and Davies (1976) employed event rates of 15 or 30 per minute in a visual task. Again d' decreased with time only at the higher event rate and was accompanied by an increase in β .

Besides the tendency toward a decrement in measured sensitivity, an increase in event rate produces a general deterioration of performance in

vigilance tasks (Jerison, 1965; Jerison, Pickett & Stenson, 1965; Loeb & Binford, 1968; Taub & Osborne, 1968; Colquhoun, 1969; Guralnick, 1973; Parasuraman & Davies, 1976). Jerison and his co-workers (Jerison, Pickett & Stenson, 1965; Jerison, 1967, 1970) have attributed this effect to the increased importance of certain modes of attention on performance at high event rates. Specifically, blurred observations of incoming stimuli could produce increased variances in the distributions of both signal and nonsignal stimuli, with the result that measured d' declines. Meanwhile, periods of non-attention, in which no responses are emitted, would result in an increase in β in those vigilance studies which require only signal responses to be made.

According to this approach, a sensitivity decrement at high event rates would result from increases in blurred observing, that is, a degradation in the quality of sensory information. Any accompanying increment in β would then reflect increases in the frequency of periods of nonattention. This would be expected mainly in visual tasks since they are loosely "coupled", in the terminology of Loeb and Binford (1964, 1968), and allow the observer to pay attention to sources of information other than the display on which signals are presented. Thus changes in both signal detection measures would be predicted for visual tasks. Auditory tasks, on the other hand, are closely coupled in that diversion of attention from the signal source is less likely. In consequence, only an effect analogous to blurred observing might be expected, in which case there should be a decrement only in d' .

In contrast, however, the opposite effect was found by Loeb and Binford (1968) in their auditory task: β increased with time at the higher event rates while d' remained constant. It would thus appear that the observed sensitivity decrements are the reflection of an effect such as "fatigue" (Welford, 1968), specific to the visual system when it is placed under a greater information-processing load. Whatever their cause, however,

sensitivity decrements are confined to a narrow range of vigilance tasks, whereas increments in β occur irrespective of the attentional demands of the task, or the sensory modality involved. An increased strictness in the criterion which must be exceeded in order to make a signal report can therefore be regarded as a major component of vigilance decrement.

(c) Interpretation of β increment by the 'ideal observer' hypothesis

Given the general finding of an increase in β with time on task it is then necessary to specify some influence on the observer's decision-making process which could bring about this increment. As Swets (1973) pointed out it could be due to a decreasing expectancy of signal occurrence, so that there is a lowered subjective estimate of $p(s)$, or to a motivational change, such as might follow a decrease in the value of a correct detection as compared to the cost of a false alarm (equation 2). The former alternative is the basis of the 'ideal observer' hypothesis as applied to vigilance by Williges (1969). According to Williges the increase in criterion typically observed reflects a tendency toward more optimal decision behaviour which, in the case of a symmetrical payoff matrix, would be indicated by values of β specified by equation 3. On this view, it is assumed that observers regard signals and nonsignals to be about equally likely at the outset of a task, corresponding to an "equal-odds" β of 1.0. Subsequent changes in β will be in accordance with the actual signal probabilities employed. For example, in the case of the low signal probability characteristic of most vigilance studies, β would be expected to increase from its initial level. Conversely, when signal probability is greater than .5 a decrease should occur.

These predictions are based on two distinct assumptions. The first is the explicit assumption that the observer usually begins a vigilance task

with the expectation that signals and nonsignals will have the same likelihood of occurrence. The second is the implicit notion that when the observer encounters a discrepancy between the actual and expected probabilities of signal occurrence, he will alter the amount of bias he has toward the signal response alternative in such a way that measured β will approach the theoretical optimum based on objective signal probability. As a consequence of this latter assumption, Williges' model can be thought of as embodying the notion of response optimization.

Colquhoun and Baddeley (1964) anticipated this approach to a great extent when they tested observers at signal probabilities that differed from those at which they were trained in order to determine whether a large part of vigilance decrement could be attributed to the frequent practice of training observers at a probability (usually .5) which was higher than that at which they were tested (usually much less than .5). In agreement with their notion of a "falsely-high expectancy" observers who were trained at a signal probability of .18 detected more signals at the beginning of a session than observers trained at a signal probability of .02, irrespective of the signal probability at which they were tested. The rate of decrement was greatest for those observers who were trained at the high and tested at the low signal probability.

Since β tended to be higher for those observers trained with a low signal probability, Colquhoun and Baddeley proposed that the signal probability encountered in training influences the decision criterion employed during testing. Unfortunately, no analysis could be performed to examine within-session revision in β . However, Colquhoun and Baddeley (1967) replicated the experiment using an auditory rather than the original visual task. Again β tended to be higher among those observers who had been trained at the low signal probability. Changes in β during the session were assessed and

it was found that β increased when observers were trained at the high and tested at the low signal probability, while a decline in β occurred among observers trained at the low and tested at the high signal probability. In each case the direction in which β was revised agrees with the ideal-O hypothesis.

The hypothesis was given its first direct test by Williges (1969), who required observers to monitor under signal probabilities of either .83 or .17. When observers were given accurate instructions about the proportion of signals to be presented during the task, both groups adopted an initial value of β around 1.5, close to the "equal-odds" value, which was then revised toward its optimal level of either .21 or 4.88. The decrease in β at the high signal probability, as in the Colquhoun and Baddeley study, then represents a vigilance increment in that β changed in a direction opposite to that normally associated with vigilance decrement. In a further study, Williges (1973) found that observers who monitored under a constant signal probability of .1 showed an increase in β during a session which in fact exceeded the optimal value of 9.0, while observers who monitored at a constant probability of .5 adopted a virtually constant β of about 1.0. Observers in a third condition monitored under a signal probability of .1 which was increased to .5 midway through the session. When compared with those observers who experienced a probability of .1 throughout the session, the tendency for β to increase was counteracted by the increase in probability. Hence it appears that decrement in vigilance performance can be reduced, or even reversed, by altering the observer's decision criterion through appropriate manipulations of signal probability.

(d) Explanation of change in response latency and confidence during a vigil: the 'distance-from-criterion' hypothesis

As well as the decrease in both detection and false alarm probabilities which would accompany an increase in β , a signal detection model of vigilance can be extended to make predictions about changes in other response measures during a vigil. In signal detection studies an increase in the probability of a signal leads to a reduction in the latency of signal decisions and an increase in the latency of nonsignal decisions (Carterette, Friedman & Cosmides, 1965; Gescheider, Wright & Evans, 1968; Gescheider, Wright, Weber, Kirchner & Milligan, 1969). As an explanation Gescheider et al. (1968) advanced the 'distance-from-criterion' hypothesis which proposes that "decisions should become increasingly difficult the closer on the sensory continuum an observation is to the criterion. As a consequence decision times should increase systematically as the magnitude of an observation approaches the criterion" (p. 501).

On this view, the relative amount of evidence in favour of a decision, as measured by the difference between the value of the **likelihood ratio** at the observation point and its value at the criterion, increases with the distance between an observation and the criterion. Decisions should be quicker when they arise from observations which lie further from the criterion, since they will be facilitated by this increased evidence. The proposed relationship between decision latency and the distance of the observation point from the criterion is shown in Figure 2. If an increase in signal probability serves to lower β , signal responses should decrease in latency, since they are based on observations which now lie further from the criterion. Conversely, non-signal responses should increase in latency, since they are based on observations now lying closer to the criterion.

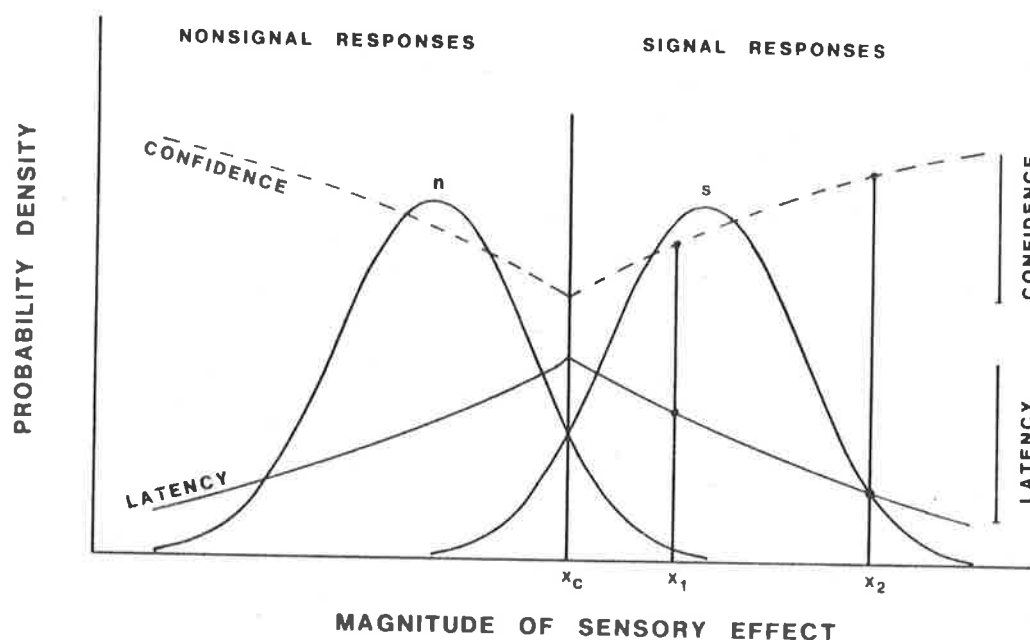


FIGURE 2. Illustration of the distance-from-criterion hypothesis as applied to both response latency and confidence. As the distance between an observation, X , and the cutoff position, X_c , is increased in either direction from X_c there is a decrease in decision latency and an increase in decision confidence. An example is from a signal detection based on X_1 to one based on X_2 .

In agreement with this hypothesis, Emmerich, Gray, Watson and Tanis (1972) found that when observers in an auditory signal detection task were required to decrease the proportion of trials on which they made a signal report (corresponding to a change from a lax to a strict criterion) signal response latency increased, and nonsignal response latency decreased. Moreover, incorrect responses had longer latencies than correct responses for both signal and nonsignal decisions. This would be predicted since, on average, observations giving rise to incorrect responses will lie closer to the decision criterion, provided that it lies somewhere between the means of the signal

and nonsignal distributions. The hypothesis would then account for the increase in detection latencies during a vigil as the result of a displacement of the observer's decision criterion to the right on the sensory continuum as β increases (Figure 2).

Response confidence can also be incorporated into the model. This follows almost by definition since the movement from a lax to a strict criterion in a confidence rating procedure can be interpreted as an increase in the distance from a fixed cutoff point. On this view, decisions made with the higher degrees of certainty would then lie further from the cutoff, and should therefore be associated with shorter decision latencies. This inverse relationship between the latency of a decision and the confidence expressed in its accuracy has been found in a psychophysical task of same - different judgements (Katz, 1970).

In the detection task of Emmerich et al. (1972) the relationship was found to hold for both signal and nonsignal decisions so that it could be proposed that as the distance from the criterion increases on either side of the criterion there will be a decrease in the latency of decisions and an increase in the confidence with which they are made, as shown in Figure 2. In a vigilance task an increase in β should then be accompanied by an increase in the latencies of both detections and false alarms and a decrease in the latencies of both correct rejections and misses. At the same time, the confidence expressed in signal decisions should decrease while that expressed in nonsignal decisions should increase.

The prediction about response latencies has been verified by Parasuraman and Davies (1976). An increase in criterion, whether induced by a reduction or signal probability or the result of the passage of time on task, was accompanied by an increase in the latencies of both categories of

signal decision and a decrease in the latencies of both categories of nonsignal decisions. The latter result has already been cited as it is contrary to the notion that there is a general decline in alertness during a vigil, but is consistent with the signal detection model outlined.

The data on changes in confidence during a vigil are scanty. Where confidence ratings have been required of observers a decline in the confidence with which signals are reported during a vigil has been found (Broadbent & Gregory, 1963, 1965; Loeb & Binford, 1964; Colquhoun & Baddeley, 1967), but confidence in nonsignal responses has not been recorded. The inverse relationship between confidence level and response latency has been found by Colquhoun and Goldman (1972) who required observers to report only signal occurrences, and made no analysis of within-session trends. There is thus no evidence as to whether the relationship between criterion placement, response latency, and confidence will hold for all decision categories and under conditions of changing criterion placement in vigilance tasks.

(e) Problems for the ideal observer hypothesis

Despite its initial plausibility, there have been some anomalous results for the ideal observer hypothesis. For example, Williges (1969) gave misleading practice and instructions to a number of his observers so that they would expect signals and nonsignals to be equiprobable, and then tested them at either of the two signal probabilities, .83 or .17. If response optimization had taken place, the instructions and practice would have led to values of β which started at an intermediate level and then approached the optimal values specified by the actual signal probabilities employed in the task. However, Figure 3 shows this to be the case only with observers who had been given accurate instructions prior to the task. Observers who had been given the

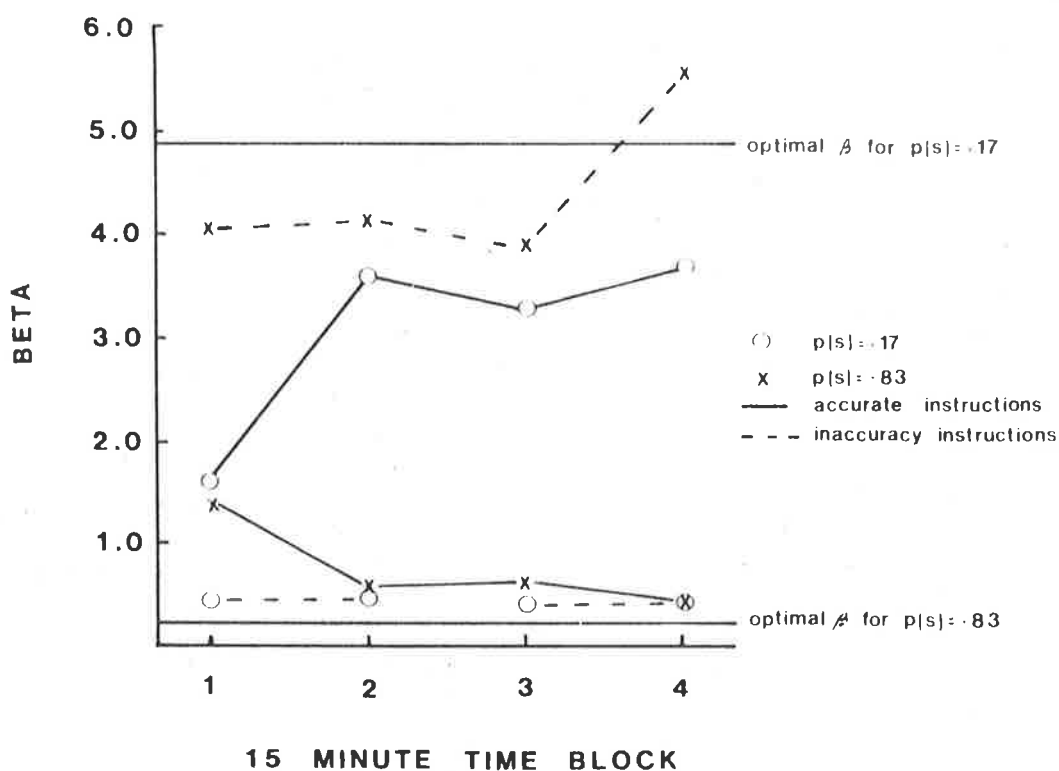


FIGURE 3. Results from Williges (1969). Values of β during the session are shown as a function of signal probability and accuracy of instructions. The optimal values of β specified by signal detection theory at the two levels of signal probability are indicated by straight horizontal lines. (Adapted from Figure 3 of Williges, 1969).

misleading instructions showed little or no revision in β during the testing session, and produced empirical values of close to 4.88 and .21, compared with ideal values of .21 and 4.88, respectively.

Williges argued that these inappropriate values of β followed from a tendency for observers to maintain the response ratios they used in training during the task which followed. As a result, observers made more signal reports at the low signal probability and fewer at the high, in order to respond with the same intermediate ratio used in training, thus producing values of β

which were inversely related to the optimum ones. Nevertheless, it seems that the adoption of a decision criterion corresponding to an intermediate value of β would be more appropriate for observers who expect signals and nonsignals to be equally likely than for observers who are led to expect, and encounter, a high or low signal probability. That is, an equal-odds β around 1.0 would have been "optimal" at the outset of the task for the observers given misleading instructions, and a higher or lower value of β for those observers given accurate instructions.

Further departures from optimality in the revision of β occurred in Williges' (1973) study in which signal probability changed midway through the session for one group of observers. These observers had been trained at a signal probability of .1 and told that it would be .5 "at the end of the session". If the goal of the observers had been the optimization of β , an initial β of 9.0 would have been adopted which then declined towards 1.0 after the change in signal probability. The actual result was an almost invariant β , initially less than 2.0 and finally slightly greater than 2.0. Initial β was then closer to the optimal value specified by the subsequent signal probability, and when the expected change in signal probability took place β failed to exhibit the decrease which was warranted. The maintenance of response ratios cannot account for the invariance of β in this task, since the increase in signal probability, combined with a stable criterion position, would have led to an increase in the relative proportion of signal reports. Conversely, a tendency for the observer to maintain an intermediate response criterion, when a change in his response ratio indicates that signal probability has changed contradicts the response optimization assumption of the ideal observer hypothesis.

Thus, it appears from Williges' (1973) study that a change in signal

probability during a task may fail to elicit a change in β in the optimal direction, at least in the case where observers are forewarned about the change. Also, Williges' (1969) study indicates that a change in signal probability from training to testing may produce values of β that are quite inappropriate at the beginning of a session and tend to remain that way throughout the session. This latter result is especially difficult to interpret since Colquhoun and Baddeley (1967) found that β changed in the direction predicted by the ideal observer hypothesis when observers were trained at a low and tested at a high signal probability, and vice versa. However, Williges (1969) found such revision to occur only if observers were trained at the same probability as that at which they are tested.

C. SOME GENERAL CONCLUSIONS

All of the theories of vigilance discussed in this chapter evolved from the common goal of providing an explanation for the decline in detection performance usually assumed to be a defining characteristic of vigilance tasks. The early learning theories of vigilance behaviour proved to be inadequate when the assumptions they made in order to account for vigilance decrement were contradicted by research into the effects on the overall level of performance of variations in signal characteristics such as frequency and time of occurrence. Of the remaining three theories presented - expectancy, arousal, and signal detection theory - the latter two appear to provide the most comprehensive account of vigilance behaviour, especially since expectancy theory could be incorporated into either of the other two theories. For instance, it could be argued that expectancy for a signal should affect the overall readiness of the observer to make a response, thereby amalgamating expectancy with arousal theory. Alternatively,

expectancy could be seen as changing the bias the observer has toward making a signal response, thereby becoming an adjunct to signal detection theory. In both cases an increase in the level of expectancy would lead to an increase in the probability of detecting a signal.

Fortunately, for the purpose of making comparisons, arousal and signal detection theories differ markedly in their interpretations of vigilance decrement. The former approach views the repetitive nature of a vigilance task as being conducive to a decline in the general level of activation of the observer, whereas the latter attributes decrement either to a change in the bias toward reporting the presence of signals or to a change in the capacity to distinguish between signal and nonsignal events. Since the general finding has been an increase in β with time on task, at least in those tasks where fatigue of the perceptual system involved in the discrimination process does not appear to be a critical factor, a contrast can be made between the arousal theory interpretation of an increment in β , and the interpretation given by the ideal observer hypothesis.

1. Arousal Versus a Response Optimization Interpretation of β Increment

It seems unquestionable that the level of arousal of the observer can determine his overall performance in a vigilance task, as well as contribute to any decline in performance which might be observed. A model, proposed by Welford (1968), predicts that a decline in arousal level should be reflected as an increase in β . According to this model, it is assumed that a decline in arousal takes the form of a general decline in neural activity, with the result that there is a reduction in the sensory effect produced by both non-signals, or noise, and signals. The distributions of noise and signal plus noise then move leftward along the sensory continuum in such a way that the

observer's discriminative ability is relatively unaffected. As shown in Figure 4, the means of the two distributions shift toward the zero-point of neural activity and, as they do so, decrease in variance so that the measure d' is not appreciably affected. A further assumption is that there is no alteration in the amount of sensory evidence required by the observer in order to report a signal, so that the position of the decision criterion remains fixed. As the distributions move to the left, the ordinate of the criterion to the nonsignal distribution will decrease, and its value to the signal distribution will increase, resulting in an increase in measured β .

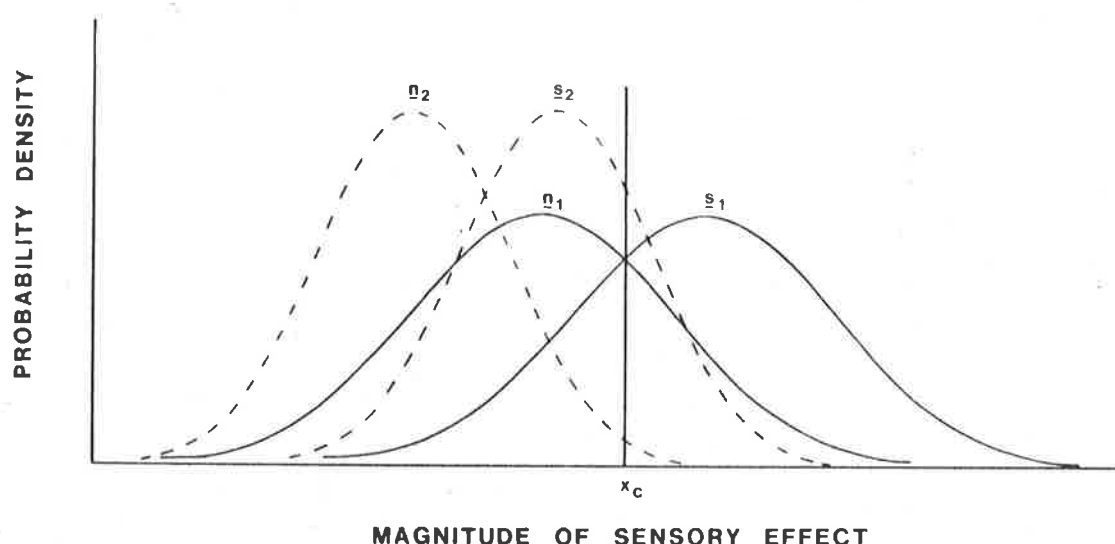


FIGURE 4. Welford's (1968) model of the effect of a change in arousal on the mean and variance of the s and n distributions. As the distributions move leftward (from s_1 to s_2 and n_1 to n_2) due to lowered arousal, they decrease in variance so that d' is little affected. If x_c remains fixed its ordinate value to s will increase and to n will decrease so that the value of β increases.

Besides lending support to arousal theory, this model demonstrates the flexibility of the basic signal detection model in the interpretation of results. It is this property, however, that has led theorists such as Jerison (1967) to question the usefulness of the underlying detection model as an explanatory theory of vigilance behaviour. [This issue will be discussed further in Chapter 5.]

While Welford's model appears to make a nice merger between two quite dissimilar approaches to vigilance decrement, its scope is limited to an explanation of performance change due to those factors which can be shown to affect some hypothetical arousal mechanism. If during the experimental session there is no change in any variable which can be assumed to influence the arousal level of the observer, then the model predicts no change in the measure β , or (at most) a gradual increase from its initial value, such as would follow from a general decline in neural activity with the passage of time. That experience of a different signal probability in a previous training session can influence the value of β during a testing session, as demonstrated by Colquhoun and Baddeley (1967) and Williges (1969), cannot be explained by the model. The level of arousal of the observer should be determined solely by the amount of stimulation afforded by the ongoing probability, and the amount of overt responding which it requires. If there is no change in this probability during a task, its effect on arousal level, and, according to Welford's model, on β , should be constant.

As shown by Williges (1969), β can either increase or decrease from its initial value, depending upon whether this value was too high or too low with respect to the signal probability employed in the session. Although an increase in β can be interpreted as being mediated by a drop in arousal, those increments and (especially) decrements, which have been shown to

depend on expected signal probability are not explicable in terms of arousal change alone. That the revision of β during a task has generally been shown to proceed in the direction of optimality, as defined by the signal probability employed during the task itself, clearly shows the potential for an interpretation of vigilance decrement in terms of a response optimization process.

2. Change in Signal Probability from Practice to Test Conditions - Problems for a Response Optimization Model

Contrary to the prediction of a response optimization process, Williges (1973) found no revision in β in response to a change in signal probability, when this change occurred during the event sequence of a vigilance task. This result may have been due to the instructions which led the observers to anticipate the probability change, so that any 'innate' response to a change in signal probability was suppressed. However, further evidence against response optimization stems from the initial values of β adopted by observers in Williges' (1969) study. This study indicated that the intermediate values of β which the ideal observer hypothesis assumes to be adopted at the beginning of a task can be clearly inappropriate to the actual signal probability with which observers had experience in training, when no change in signal probability occurs from practice to testing conditions. When a change does occur from training to testing, the values of β are also inappropriate to the actual signal probability used in testing. In neither case has the observer adopted a decision criterion which results in an optimal value of β .

The actual adaptation which follows a change in signal probability cannot be studied in the usual experimental paradigm, where the change occurs from practice to testing conditions, thus preventing an ongoing trial-by-trial analysis. Hence, in the initial experiments to be reported in the

next chapter, signal probability was decreased in a gradual fashion throughout the task from an initial value of .5. A tendency toward response optimization should be observed as a systematic decrease in β from an initial equal-odds value toward those values specified by the actual presentation probabilities.

A subsidiary aim of the study was to observe whether the increase in the criterion would be accompanied by a decrease in both the speed and confidence with which signals were reported, and an increase in the speed and confidence of nonsignal reports, as implied by the proposed "distance-from-criterion" hypothesis.

CHAPTER 2

The apparent advantage of the signal detection model of vigilance, when combined with its auxiliary ideal observer and distance-from-criterion hypotheses, lies in its potential ability to account for the effects of a number of independent variables, as well as the framework it provides to evaluate the relationship between response variables not normally measured in vigilance tasks. For instance, while arousal theory can successfully explain the decline in the speed of signal responses during a vigil, it cannot explain the concomitant increase in the speed of nonsignal responses, which went unmeasured until the study of Parasuraman and Davies (1976). Nor can arousal theory, in its present form, give a satisfactory explanation for the confidence with which observers make their decisions except, perhaps, by arguing that the speed of responding acts as a cue by which the observer gauges his confidence.

The two studies reported in this chapter were focussed on the unfounded response of the observer to a change in ongoing signal probability. They therefore required a task in which the change would take place in such a manner that the amount of incoming stimulation from the task, as well as the amount of overt responding it required, would remain constant. If this were accomplished, only the observer's decision criterion, and not his overall level of arousal, would be affected. The task selected was one in which there was a change in the relative frequency of two alternative events, both of which required an overt response, and either of which could be categorized as 'signal'. Hence, in order to evaluate whether those changes in the decision criterion which followed a change in signal probability were in fact those specified by a signal detection model of vigilance, it was necessary to employ a task which differed somewhat from the traditional vigilance task.

A. EXPERIMENT ONE : EFFECTS OF A STEADY DECREASE IN SIGNAL PROBABILITY

A discrimination task that required a judgment as to which was the longer of two vertical lines was chosen for the present study, as it should not give rise to any a priori disposition to either response alternative. By comparison, in the traditional vigilance task, more importance has been attached to the detection of signals (which are often defined as "critical events") than to their correct rejection. If the observer attempts to maximize the expected value of a decision, therefore, an unduly low value of β may result. Observers show the same shifts in criterion position when the value of making a correct signal response is increased, as when a priori signal probability is increased (see Green & Swets, 1966, Chapter 4). Hence, criterial shifts due to a change in signal probability might be confounded with, or diminished by, shifts toward a more symmetrical payoff matrix due to some motivational change of the observer during testing.

A discrimination task which requires continuous monitoring, can be likened to a vigilance task in which equal weightings are given to the values of correct detections and rejections, and to the costs of false alarms and misses. Either of the two physical events, the longer line to the left or the longer line to the right, could be arbitrarily designated as the signal event for analytical purposes, although never defined that way in the presence of the observer. If the two events are initially equiprobable, and their physical magnitude is held constant, subsequent variations in their relative probabilities should lead to performance changes that reflect the movement of β away from an initial equal-odds value toward the optimal value as specified by equation 3 in Chapter 1.

The design of the task took the form of a constant decline in the

relative probability of one of the events, the 'signal', which according to the ideal observer hypothesis should give rise to a steady increase in the criterion level for a signal report. Progressively fewer observations will exceed the new criterion level, and those which do will now lie closer to the criterion. According to the distance-from-criterion hypothesis, therefore, there should be a decline in both the speed and confidence with which signal reports are made. Meanwhile, complementary changes should occur in the pattern of nonsignal responses, which should be reported with increased probability, speed and confidence.

1. Method

(a) Task and stimuli

On each trial the observer had to judge which was the longer of two vertical lines which formed the lower segment of a π -figure stimulus presented on the screen of a television monitor as shown in Figure 5. The longer line was always 141 mm. in length and the shorter was 140 mm., and they were separated by a distance of 40 mm. The lines were viewed at a distance of 1.6 m. so that the difference in line lengths subtended a visual angle of 0.02 degrees. Each decision response was followed by a rating of confidence in its accuracy along a 10-point scale. The stimulus was visible for 0.2 seconds and total trial length was 6 seconds. The decision response had to be made within 3 seconds of stimulus onset and the associated confidence response had to be made before the next stimulus appeared. A response not made within these periods was classified as a "no decision".

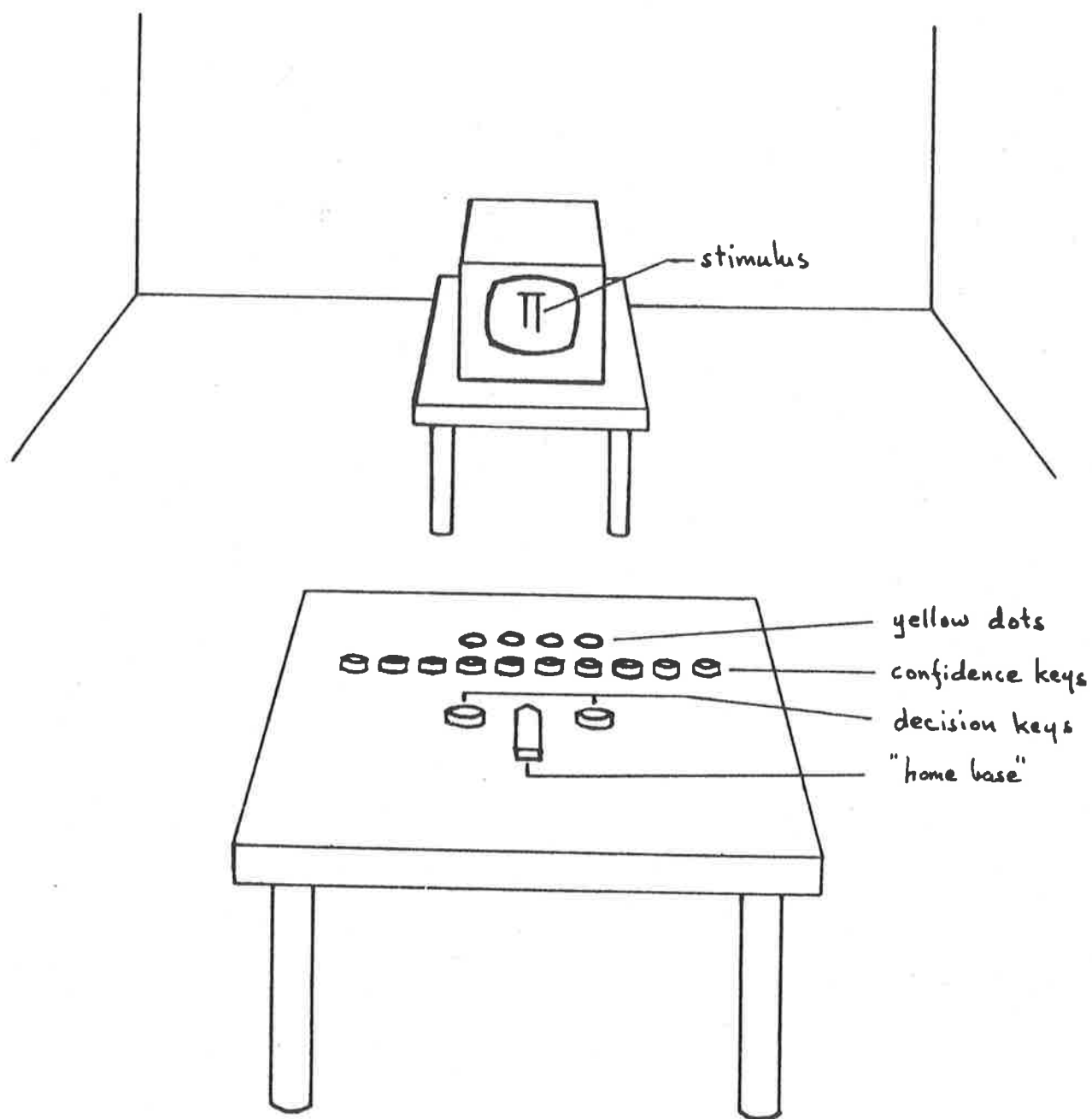


FIGURE 5. Response and stimulus presentation apparatus as used, with only minor modifications, throughout Experiments 1 to 7. Presented on the television monitor in the above example is a π -figure stimulus of the type used in Experiments 1, 2, 3 and 6.

(b) Apparatus

Stimuli were generated by means of a graphics program executed by a D.E.C. PDP 8/F computer. The side on which the longer line appeared was determined by a disk-based PDP 8/E computer which also regulated stimulus duration and inter-trial interval. This second computer also categorized each primary decision, measured its response time and confidence, and provided print-outs of summaries of these statistics.

Observers depressed either the left or right of two perspex buttons, lying on either side of a wedge-shaped resting position, to indicate whether the left or right line, respectively, was judged longer. Following this response, confidence was expressed by depressing one of a row of ten confidence buttons lying in front of the primary response buttons (Figure 5). The confidence scale was arranged from "least confident" at the leftmost button to "most confident" at the far right.

(c) Design

An experimental session consisted of 25 demonstration trials, two practice blocks of 50 trials each, and a main session of 600 test trials. During the demonstration trials the longer line occurred to the right and to the left in an alternating fashion. In the two practice blocks stimuli were presented randomly with the constraint that the longer line be equiprobable on either side. The main session contained six blocks of 100 trials each, with no break between blocks so that it appeared as a continuous sequence of stimuli. The stimulus with the longer line to the right was arbitrarily designated as the signal. The relative probabilities of the signal stimulus for the six consecutive experimental blocks were .50, .41, .32, .23, .14, and .05, respectively. Due to the random nature of the sequence in which stimuli were generated,

there was a progressive and fairly continuous decline in the probability of the signal during the course of the task. Meanwhile, the nonsignal stimulus increased in probability in an exactly complementary fashion from .50 to .95.

(d) Procedure

Observers were tested individually and in complete isolation with watches being removed before subjects entered the experimental chamber. Instructions were given verbally by the experimenter. Observers were told to respond to the longer of the two lines as quickly as possible without making **too many mistakes and, during practice, to express confidence in that decision by** depressing one of the four central confidence buttons marked with yellow dots (Figure 5). Observers were instructed to spread their responses over all four buttons and to develop a steady rhythm of making a decision response, followed by an expression of confidence in that response, before finally returning the finger to the 'home base' before the start of the next trial. All responses were made with the index finger of the dominant hand.

Observers were told that during the demonstration and practice trials the longer line would occur with the same frequency on each side. The demonstration trials were followed by the two blocks of practice trials. A break of approximately three minutes occurred after each block during which data for that block was printed out by the PDP 8/E and a check made to determine whether the observer was responding appropriately. The demonstration and practice trials, together with the 60 minute main session, occupied a total time of about 80 minutes.

Before the main test session observers were given the following additional set of instructions:

"You may sometimes feel that the relative frequency with which the left or right is longer has changed, that is, that the longer line is occurring more often on one side than on the other. Don't worry, since it may well be the case. Also, remember that the bottom of the range of confidence keys represents the least confident response you made in the practice trials, while the top, right-hand side, represents the most confident response you made in the practice trials. Keep using these keys to represent those levels of confidence. However, should you feel even more, or even less, confident please use the additional keys above and below the yellow range. Use keys further to the left if you feel less confident, and further to the right if you feel more confident. "

These additional instructions were intended to prevent the observer from responding under the mistaken assumption that relative probabilities were to remain constant. Secondly, the instructions to extend the range of confidence responses were meant to enable a more sensitive measure of possible confidence change as a function of experience with the task. It was felt that if observers were restricted to the range of confidence used during practice trials, then these measures might exhibit an artificial "ceiling". This effect would be particularly accentuated among observers who tended to use the upper, or lower, limits of the confidence range for any category of response during practice.

It became apparent after a few observers had been tested that there was a bias toward making the "right longer" response. The majority of incorrect responses during practice and the first experimental block, where relative probabilities were .50, were due to incorrect reports of the stimulus with the longer line to the right. Consequently, half the observers were tested with the stimulus having the longer line to the right as the signal, and half with the stimulus having the longer line to the left.

(e) Observers

Twenty-six female undergraduates enrolled in the second year Psychology course at the University of Adelaide volunteered to participate at the rate of \$2.75 per hour. Ages ranged from 18 to 39 years, with a median age of 19 years. All observers had normal or corrected vision.

2. Results

A second check of the equipment used in the experiment was carried out after the last observer was tested. Unfortunately, this revealed that occasionally and unpredictably confidence responses were being incorrectly recorded by the system. Even after exhaustive investigation it was not possible to determine whether this arose from some fault of the hardware or from some inadequacy in the overall system program, and even a trial-by-trial analysis of data could not reveal where mistakes had occurred in the data which had been recorded. As a result no reliance could be placed on the confidence data and the following discussion is therefore restricted to the measures of response probability and response times.

(a) Response probability

(i) General trends. Changes during the session in the conditional probabilities of correct and incorrect signal reports are presented in Figure 6. The trends which accompanied the shifts in relative stimulus probabilities run contrary to the predictions of the ideal observer hypothesis. The probability of correctly reporting a signal appeared to increase as a function of the number of trials, as did the probability of incorrectly reporting a signal on nonsignal trials.

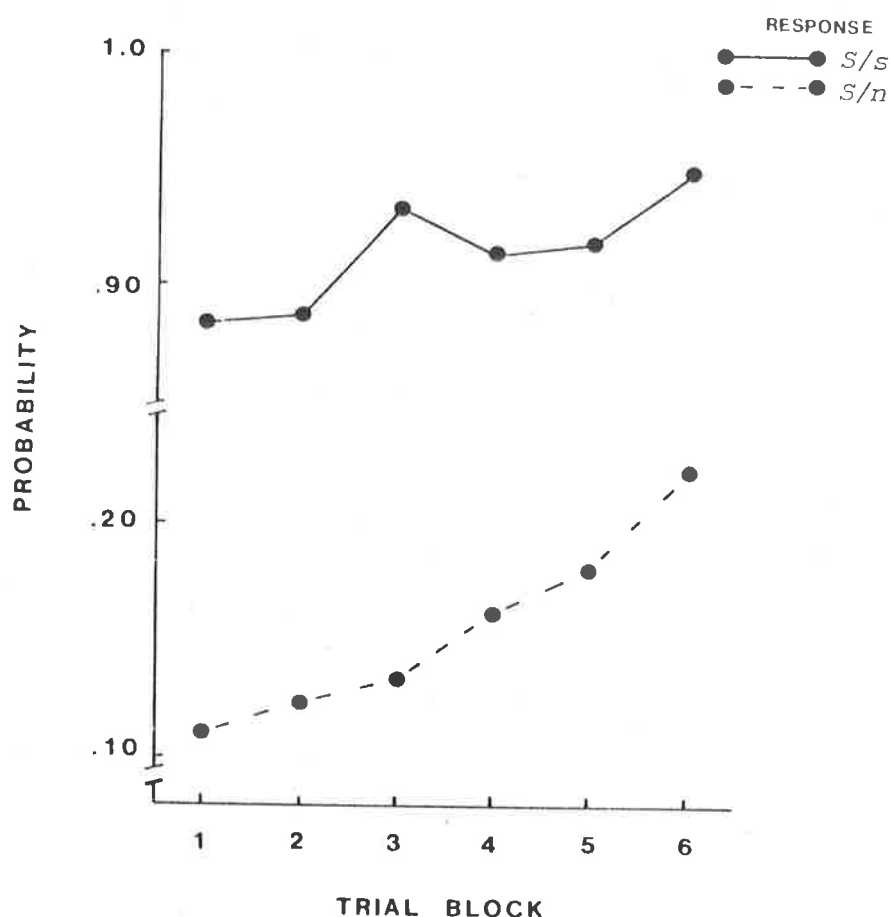


FIGURE 6. Changes in the mean conditional probability of a correct (S/s) and incorrect (S/n) signal response over the six trial blocks.

(ii) Effects of side. In order to examine the trends in more detail, data for each observer were averaged over the initial and final three experimental blocks. Analyses of variance were then performed on the probabilities of making a signal response correctly and incorrectly, according to the main factors of Side (signal on left or signal on right) and Period (first or second half of session). Results of these analyses are given in Table 1.

As can be seen from Table 1, there was a strong effect on both responses according to whether the 'right-longer' or 'left-longer' stimulus was designated 'signal'. Specifically, the probability of a correct signal decision was higher (.97) when the 'right-longer' stimulus was designated as the signal

TABLE 1. Results of analyses of variance on conditional response probabilities.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
<u>(a) Correct signal</u>					
<u>response P(S/s)</u>					
	Side	1	.108	18.770	.0005
	S within Side	24	.005		
	Period	1	.009	3.257	
	Side x Period	1	.002	.942	
	Period x S				
	within Side	24	.002		
<u>(b) Incorrect signal</u>					
<u>response P(S/n)</u>					
	Side	1	.166	11.262	.005
	S within Side	24	.014		
	Period	1	.047	24.900	.00005
	Side x Period	1	.001	.950	
	Period x S				
	within Side	24	.001		

than when the 'left-longer' stimulus was so designated (.87). Similarly, the probability of making a signal decision incorrectly was higher in the former (.20) than in the latter case (.08). This effect seems to be due to the apparent bias for reporting the right line as longer and, as mentioned above, was noticed early in the testing of observers.

(iii) Effects of period. Turning to the effect of Period, there was an increase in the conditional probability of making signal responses which was significant only for incorrect signal reports. As Figure 6 indicates, the initial increase in the probability of a correct signal decision appears to "level off" after the third trial block. However, when a comparison was made between the data of the first and sixth trial block, the increase was significant ($F = 6.86$, $df = 1/24$, $p < .05$). On the other hand, for both correct and incorrect signal responses, there was no significant interaction

between the side on which the signal occurred and period of the task.

Therefore, whether there was an initial tendency to over-respond or under-respond to a particular stimulus, as the probability of that stimulus decreased, the probability of correctly reporting it in its presence, and falsely reporting it in its absence, increased.

At the same time, the greater room for increase that was possible in the probability of incorrectly reporting a signal than was possible for correctly reporting one, appears to have led to a deterioration in overall performance. Specifically, in the probability of making a correct decision (irrespective of whether this meant a signal or nonsignal response) there was a significant decline from the first to the second half of the session ($F = 5.315$, d.f. = 1/24, $p < .05$). On the other hand, there was no difference due to the side designated as signal ($F = .55$, d.f. = 1/24, $p > .4$), nor any interaction between period and side ($F = 2.621$, d.f. = 1/24, $p > .1$). Thus the unconditional probability of making a correct decision was the same, and showed the same decrement, irrespective of whether the left-longer or right-longer stimulus was designated 'signal', and, hence, decreased in probability as a function of number of trials.

(iv) Change in 'local' and 'cumulative' probabilities. An explanation for the fact that the conditional probabilities of making a signal report changed in the opposite direction to that predicted is that response ratios were revised in the correct direction, but "lagged behind" the change in actual stimulus ratios. That is, although observers tended to over-respond to the signal as it decreased in probability, the probability of making a signal report, both correctly and incorrectly, decreased with time. Revision in the ratio of signal to nonsignal decisions thus tended to be conservative with regard to the actual change in stimulus ratios.

In Figure 7 the probability of making a signal report, conditional on either stimulus, is compared with two objective probabilities, the probability with which the signal stimulus occurred during an experimental block of 100 trials (local probability) and the probability with which the stimulus occurred when averaged over all trial blocks up to the end of that block (cumulative probability). As can be seen, the unconditional probability of making a signal report decreased linearly, and most closely followed cumulative probability rather than the actual probability of the signal during any trial block. As a test of this finding, best-fitting straight lines were calculated for signal

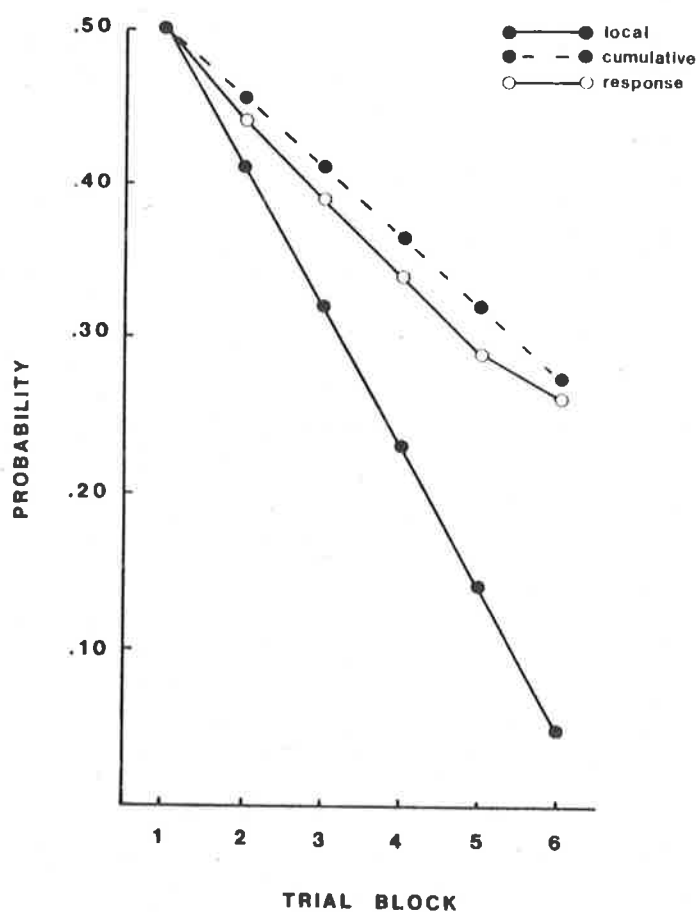


FIGURE 7. Changes in the mean unconditional probability of making a signal response compared with the local and cumulatively-based probability of a signal stimulus over the six trial blocks.

response probability plotted as a function of block probability, and as a function of cumulative probability. A perfect correspondence for either function would have been a line with a slope of 1.0 and an intercept of 0.0. The actual slopes and intercepts were .58 and 0.20, respectively, when signal response probability was plotted as a function of local probability; and 1.15 and -0.09, respectively, when plotted against cumulative probability. These latter statistics are much closer to the theoretical.

In order to statistically verify that signal response probability matched cumulatively-based signal probability more closely than the local probability of a signal, a crude type of goodness-of-fit test was performed. For each observer, the difference between the obtained slope and intercept and their respective ideal values of 1.0 and 0.0 were computed, in both the case where response probability was plotted against cumulatively-based signal probability and that where it was plotted against local signal probability. For each of these sets of differences Wilcoxon matched-pairs signed-ranks tests were performed to see whether the obtained and theoretical parameters could be considered to have arisen from different populations of slopes or intercepts. In the case of signal response probability plotted against cumulatively-based signal probability, it was found that there was no difference between the obtained and theoretical parameters for both slopes ($T = 124$, $n = 26$, $Z = -1.31$, $p = .19$ (2-tailed)) and intercepts ($T = 111$, $n = 26$, $Z = -1.64$, $p = .10$ (2-tailed)). However, in the case of signal response probability plotted against local signal probability, there was a significant departure from the theoretical value in the case of slopes ($T = 0$, $n = 26$, $Z = -4.46$, $p < .0001$ (2-tailed)) as well as intercepts ($T = 2$, $n = 26$, $Z = -4.41$, $p < .0001$ (2-tailed)). At this stage, therefore, it appears that if the observer estimates the probability of an event from his own corresponding response probability, then this estimation seems to be based on some temporal averaging process.

(b) Individual differences in the revision of response ratios

While the averaged data presented above indicates that a consistent and systematic change in response ratios followed upon change in the stimulus ratio, there were large differences among observers in the degree to which response probabilities were revised. Although the average slope for the probability of a "signal" response plotted against the objective cumulatively-based probability of a signal (slope score) was 1.15, individual slope scores ranged from as low as .18, indicating very little change in response ratio, to as high as 1.98, suggesting exaggerated adaptation.

Figure 8 shows the individual curves for the two extreme slope scores, as well as an intermediate slope score. The wide range in slope scores was not due to the bias to make "right longer" responses, which could have resulted in a greater or lesser tendency to adjust response ratios depending on whether the probability of a longer line decreased on the left or the right. The mean difference between slope scores obtained when the probability decreased on the right (mean: 1.14; range: .59, 1.97) and on the left (mean: 1.17; range: .18, 1.98) was not close to significance ($t = .16$, $df = 24$, $p > .2$). The degree to which response probabilities were revised was thus not influenced by any initial response bias.

One possibility is that differences in the degree to which response ratios were revised arose from individual tendencies toward "conservatism" or "liberalism" in probability revision. While there was no exact test for this hypothesis, it could be argued that if conservatism were an individual trait, there might be some personality measure which would correlate with degree of slope score. Fortunately, records were available which contained personality data for the majority of observers. These were compiled the previous year and consisted of scores on Catell's 16 P.F. Test (Form C) and the Eysenck Personality Inventory (E.P.I.). A Pearson correlation coefficient

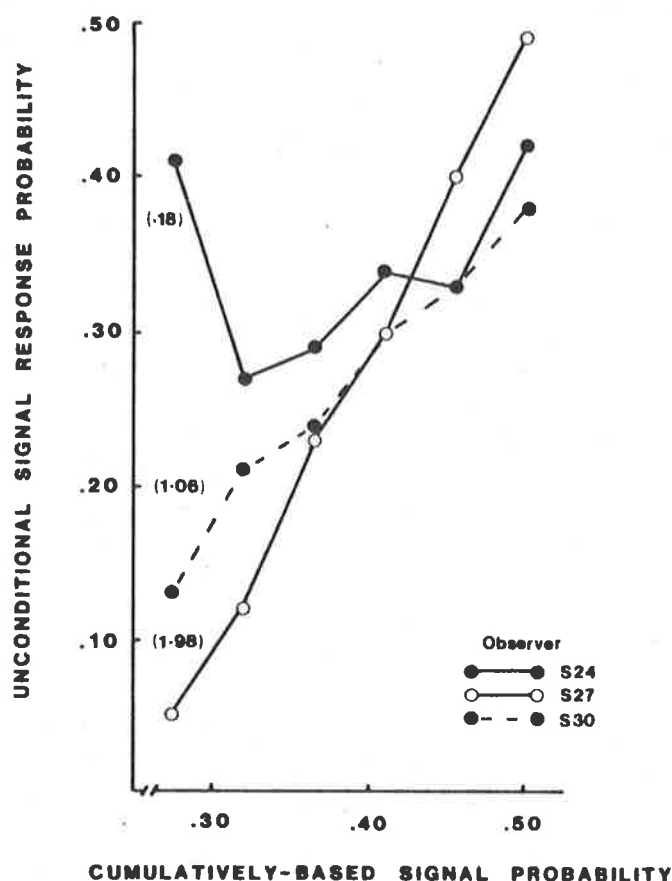


FIGURE 8. Examples of unconditional signal response probability plotted as a function of cumulatively-based signal probability. The functions were obtained from three observers, the two with the most extreme (high and low) values for the slope of the best-fitting straight line applied to each function, and one with an intermediate value. The individual slope values are given in brackets.

was therefore calculated between each personality factor and the slope score ("signal" response probability plotted against cumulatively-based objective signal probability). One correlation which did turn out to be significant was a coefficient of .46 (19 cases, $p = .02$) with the Extraversion scale of the E.P.I. Although this result should be treated cautiously, due to the high number of correlations from which it emerged, it seems a rather high correlation when it is considered that individuals were not selected on the basis of any personality measure.

(c) Signal detection analysis

Values d' and β , based on the mean probabilities for all observers in each trial block of correctly or falsely reporting a signal, are shown in Figure 9. The figure suggests that the within-session trends observed in the response probability data are the reflection of a change in the bias toward reporting a signal rather than any change in the sensitivity of observers. In order to test this possibility the response probability data were averaged over the first and final three experimental blocks in order to calculate d' and β . Results of the analysis are given in Table 2 and are based on values

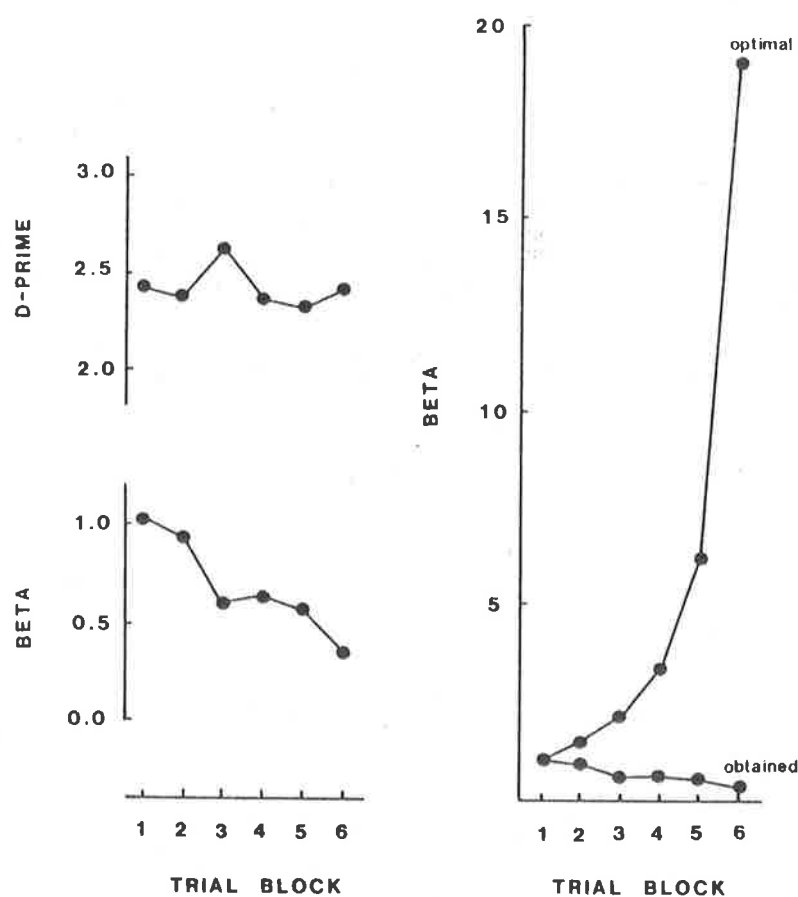


FIGURE 9. Changes in group values of d' and β over the six trial blocks. On the right the values of β that were observed are compared with the optimal values specified by signal detection theory.

TABLE 2. Results of analyses of variance on signal detection measures.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
<u>(a) Sensitivity (d')</u>					
	Side	1	.642	.705	
	S within Side	16	.910		
	Period	1	.685	3.315	
	Side x Period	1	.035	.170	
	Period x S within Side	16	.206		
<u>(b) Bias (β)</u>					
	Side	1	63.670	3.907	
	S within Side	16	16.294		
	Period	1	16.625	5.405	.05
	Side x Period	1	4.051	1.317	
	Period x S within Side	16	3.075		

of the signal detection measures which could be obtained from seven observers who had the signal on the right and eleven who had the signal on the left.

Values could not be computed for the remaining observers because of their almost complete accuracy in reporting the signal stimuli in the last half of the session. However, group values of d' and β for those individuals excluded from the analysis were 2.92 and .39, respectively, for the first half of the session. In the second half, d' increased slightly to 3.03 and β declined to .16 among the same observers. These trends support the lack of significance of the watch period effect on d' , and the significance of the same effect on β , shown among the observers whose detection measures could be analyzed. Thus, there was no change in sensitivity corresponding to the decline in the probability of making a correct decision. The tendency to over-respond both correctly and incorrectly, which caused this decline, was accompanied only by a fall in β , from an initial optimal value of 1.0 to a final value of .35.

In Table 2 there is some substantiation for the proposed existence of a bias toward making "right longer" judgments. The effect of side on d'

was not close to significance ($F < 1.0$), whereas there was a marginally significant effect of side on β ($p < .10$). At the same time there was no interaction effect, so that β tended to decrease whether it was initially high or low. Specifically, when the signal occurred on the right, β dropped from .38 in the first trial block to .12 in the last, and from 2.14 to .63 when the signal was on the left.

(d) Response latencies

TABLE 3. Mean latencies for the four categories of response (in milliseconds).

<u>Signal responses</u>		<u>Nonsignal responses</u>	
Correct (S/s)	Incorrect (S/n)	Correct (N/n)	Incorrect (N/s)
1352	1578	1414	1492

(i) Differences among the four response categories. Mean latencies for all four response categories are given in Table 3. Each possible pair of mean differences was subjected to a paired-samples t-test, and all six differences were found to be significant. Of interest are the following three results: (1) both correct signal and correct nonsignal reports took less time than their corresponding incorrect reports ($p < .001$ in each case), (2) correct signal reports took less time than correct nonsignal reports ($p = .002$), and (3) incorrect signal reports took more time than incorrect nonsignal reports ($p = .012$).

The first finding is consistent with the distance-from-criterion interpretation of response latency, since responses giving rise to incorrect responses lie, on average, closer to the criterion and so should result in the

longer decision time (see Figure 2 in Chapter 1). The second is also in agreement if, as already has been shown, β is less than 1.0, in which case the criterion is closer to the mean of the nonsignal distribution and observations which gave rise to correct signal reports lie further from the criterion. However, this also implies that incorrect signal reports should take less time than incorrect nonsignal reports. The additional assumption that the variance of the signal distribution is greater than that of the nonsignal distribution is then needed to explain the third result.

The configuration shown in Figure 10 is the only one which would result in incorrect signal reports taking more time than incorrect nonsignal reports when response bias is in favour of the signal response. Such a configuration could arise if the decrease in signal probability had led to an

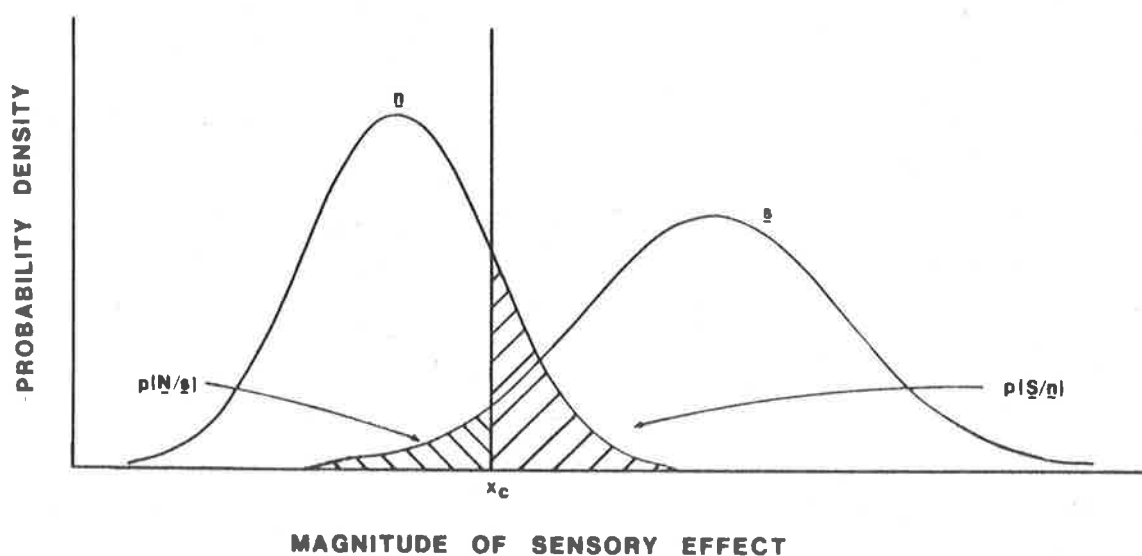


FIGURE 10. Signal, s , and nonsignal, n , distributions with variances such that values of X leading to incorrect signal reports (S/n) are distributed more closely to the cutoff value, x_c , than values of X leading to incorrect nonsignal reports (N/s). The placement of the decision cutoff in the figure corresponds to a value of β less than unity.

increase in the variance of the signal relative to that of the nonsignal distribution. However, such an effect should have led to a decrease in d' during the session, and this was not found.

(ii) Within-session trends. Throughout the session there was a tendency for response latency to change only in the case of correct nonsignal reports, which increased from a mean of 1379 milliseconds during the first half of the session to 1449 milliseconds over the last half ($F = 8.284$, $d.f. = 1/24$, $p < .005$). On the other hand, analyses of variance on the latencies of the other three decision categories showed no significant effect ($p > .10$ in all cases), and no trends were observable in the data. The increase in the latency of correct nonsignal reports could be interpreted as the result of a decreasing criterion, but some change should then have occurred in the other latency measures as well, and an increase in latency due to fatigue with time on task is a possible explanation. There is therefore dubious support for a distance-from-criterion hypothesis.

3. Discussion

(a) Adaptive response to a change in signal probability - comparisons with previous findings

The concern of this present study was the effect of an objective drop in signal probability from an initial level of 0.5, which has frequently been considered as the level of "subjective expectancy" at the outset of a vigilance task. The decrease from this level toward the actual signal probability can be regarded as similar to the initial 'adaptation' stage in vigilance between practice and test trials. The results obtained, however, directly oppose the ideal observer hypothesis. As the probability of a signal declined there was

an increase in the probability of reporting a signal, both correctly and incorrectly, and a steady decline in the criterion.

This decline in β appears to be due to a tendency for observers to respond at a rate which mirrors the cumulatively-based probability of a signal, rather than its local probability during any block of trials. Thus the tendency to over-respond to the signal progressively increased during the course of the task. This can be compared with the results found by Williges (1969) when individuals were trained to expect an equal likelihood of signal and nonsignal events and were subsequently tested at a high or low ratio of signals. Williges found that measured β was low for the low probability of signal occurrence and high for the high probability, which he argued was due to a tendency for observers to respond with those response ratios adopted during training. There was thus under-responding to a high signal probability and over-responding to a low signal probability. In Williges' (1973) study the failure of observers to adopt a lower criterion when signal probability increased midway through the task can again be interpreted as a case of under-responding to a high signal probability.

An important difference in the design of the present task, when compared with previous studies in which a change in signal probability has been studied, is that it allowed an analysis of the adaptation to a decreasing, rather than a decreased, signal probability. That is, the usual abrupt drop in signal probability encountered at the beginning of a task was extended over the entire course of the present one. Although Williges (1969) also found a lowered β in response to a drop in signal probability, he found no change in β with time on task. This implies a very rapid adaptation at the beginning of a task when observers are first exposed to the change in signal probability. The concept of 'rapid adaptation' to an altered signal probability

goes against the interpretation of vigilance decrement as a rather long-term adjustment in an initially falsely-high subjective probability. It does, however, receive empirical support from studies by Mackworth (1950, Clock Test 9), Kappauf and Powe (1950) and Jerison (1963), which indicate that the usual view of vigilance decrement as being some long-term process might be the result of too coarse a time-grain analysis of performance. Trial-by-trial analyses in these studies revealed that signal detection probability dropped and asymptoted very rapidly at the beginning of a session. A coarse analysis (for example, a comparison of performance in the first and second halves of the session) would have masked the rapid change at the beginning of the task.

The design of the present experiment caused the initial adaptation to a change in signal probability, which normally occurs at the beginning of a task, to continue throughout its entire course. If observers display response maintenance when they encounter a change in signal probability, as Williges (1969) argues, there would have been no change in the ratio of signal to non-signal reports during the present task. That this was not the case was evidenced by the significant change in the unconditional probability of a signal report during the task. Further, the revision in response ratios proceeded in the correct direction, but lagged behind the objective stimulus ratios. This revision in response ratios was accompanied by a revision of the decision criterion in a direction opposite to that specified by the optimal value of β . However, this change in the placement of the decision criterion can be regarded as being optimal if the observer's goal is to respond with a signal to nonsignal ratio that is compatible with the cumulatively-based, rather than local, probability of signal occurrence.

(b) Individual differences in the revision of response ratios

The line of reasoning presented in the previous paragraph might be capable of explaining the significant correlation between the rate at which signal response probability declined and the degree of Extraversion of the observer. In vigilance studies Introverts exhibit less detection decrement with time on task (e.g. Bakan, Belton & Toth, 1963; Keister & McLaughlin, 1972). Introverts would then be, on the present view, the poorer adapters, as initial response probabilities must be maintained in order to prevent an increase in β and subsequent detection decrement. In the same way, the Introverts of the present study showed the greatest tendency to maintain their initial response probabilities, and followed changes in objective signal probability to a lesser degree than Extraverts.

(c) The 'target' of revision of response ratios - general conservatism viewed as purposeful probability matching

The actual manner in which the probability of making a signal response declined can best be analyzed within the framework provided by studies of probability estimation, learning, and revision. In the first type of task, Underwood (1969) argued that the frequency attribute of an event exists as one of seven distinct memory components, and that the ability to estimate event frequency arises from the ability to judge the relative frequency, or probability, of events. Although relative frequency estimation is a highly developed skill, a general conservatism appears to prevail. For example, Attneave (1953) found that in judgments of the relative frequency with which alphabetic letters occurred in the English language, individuals were very accurate in their ranking of letters. However, errors of a "lawful nature" were made: low frequencies were systematically over-estimated and high

frequencies under-estimated. Although concerned with the estimation of existing internal probabilities, this result is analagous to the signal over-estimation, and nonsignal underestimation, found in each phase of the present study.

Of greater relevance are tasks in which previously unknown probabilities must be estimated and which might more directly reveal the actual process by which probabilities are learned. Laboratory studies of probability learning have usually taken the form of the "prediction task" in which an individual predicts which of two events will occur on the next trial. The prediction must thus be based on knowledge of the relative probabilities of the two events in the preceding sequence of trials. For example, there seems to be good recall of the preceding event schedule even when fewer of the dominant events are predicted than the actual schedule would warrant (Vitz & Hazan, 1969). Hence, although the individual's prediction behaviour would support an interpretation of conservatism, his memory of the preceding sequence is accurate and not conservative. Further, a short-term, local change in the relative frequency of events causes a change in the correct direction in the proportion of predictions, and this is superimposed on the more gradual change in cumulatively-based probability (Edwards, 1961).

The parallel with the results of the present study is obvious: subjective expectancy in a prediction task would also start at a .5 level and, as in the present results, the response ratio, from which subjective probability is inferred, is corrected slowly even though memory for the event sequence is accurate. Some higher-order mechanism must therefore be exerting constraint on the predictor, or observer, which causes his behaviour to be "out of phase" with his memory. Conservatism would seem to imply a lag in both memory and behaviour, and so is insufficient as an explanation.

The mechanism imposing constraints might be the "simple cumulative mechanism" which Howell (1970) formulated in order to explain the storage of relative stimulus frequency information. Howell required individuals to estimate the relative proportions of two events. Seven blocks of 20 trials each were presented, and estimations were required for either the first block, and the last block, or the entire sequence. Relative proportions for the first and final block could be equal to, or differ from, the overall value for the entire sequence. It was found that judgments made on the first block reflected the actual proportions for that block, while judgments made after the final block reflected the overall proportion for the entire sequence, regardless of the actual proportions of the block itself. This result corresponds to the present finding that response probability reflected the cumulatively-based probability for the whole preceding sequence rather than the local probability in the preceding group of trials. It also agrees with an earlier study by Robinson (1964) in which individuals were required to track changes in the relative proportion of two flashing lights with an apparatus which indicated probability estimates. Individuals followed changes in the relative proportion of the two events, and tended to successively approach, rather than over-shoot, the new probability. Hence the degree to which probabilities are revised, as well as estimated, seems to be influenced by some cumulative process.

Another type of study, in which subjective probability revision is compared to ideal Bayesian revision, is also pertinent. This is usually an investigation of the degree to which individuals revise the probability with which they hold one or two alternative hypotheses to be true, as a consequence of evidence favouring one or the other hypothesis. The amount of subjective revision is then compared with the ideal amount as specified by

Bayes' Theorem. While procedurally quite different from estimation and prediction tasks, and concerned with a different sort of subjective probability, the results which have been obtained in this type of study show a clear parallel in interpretation, i.e., a tendency for subjects to display general conservatism, subjective revision is in the correct direction but its magnitude is less than would be predicted by Bayes' Theorem, whether revision occurs after a single datum of evidence (e.g. Peterson & Miller, 1965; Peterson, Ulehla, Miller, Bourne & Stilson, 1965), a sample of data (e.g. Peterson, Schneider & Miller, 1965; Phillips & Edwards, 1966; Pitz, Downing & Reinhold, 1967), or when the sample data first favours one hypothesis and subsequently favours the alternative (Peterson & DuCharme, 1967). It also appears to be independent of the particular mode of response used to measure, or infer, subjective probability (Beach & Wise, 1969; DuCharme & Donnell, 1973; Beach, 1974).

Although theorists such as Phillips and Edwards (1966) have considered the pervasiveness of conservatism in these tasks to reflect a "limitation" on the ability to process information, and although DuCharme (1970) has interpreted it simply as the reflection of a bias against the use of extreme probabilities, the conservatism evidenced in these tasks could also be the result of a mechanism similar to that which appears to function in probability learning and estimation tasks. For instance, a sample of data includes evidence both confirming and disconfirming for a particular hypothesis. Given that a cumulative mechanism was influencing the level of subjective probability, probability changes in opposing directions would be "averaged out". The difference between subjective and Bayesian revision would then be greater, or even more "conservative", following longer sequences of data where the averaging process would have had its greatest effect since more individual

probabilities, one for each datum presented, would have been averaged together. In agreement with this view, degree of conservatism in subjective revision has been shown to be directly proportional to sample size (Peterson, Schneider & Miller, 1965; Pitz et al., 1967).

(d) Individual differences in the processing of probabilistic information
- the need of a further study

Probability revision studies are also relevant to the individual differences found in the degree to which response ratios were revised following a change in objective probability. In the study by Peterson, Schneider and Miller (1965), individual accuracy measures of probability revision on one sample size of data correlated significantly with measures on other sample sizes. Although subjective revision generally lagged behind Bayesian revision, some individuals showed consistently more lag than others. Peterson, Ulehla, et al. (1965) examined two characteristics of subjective probabilities; veridicality, the extent to which the value of a subjective probability agreed with the corresponding objective probability, and internal consistency, the extent to which a set of subjective probabilities conformed to the implications of mathematical probability theory. Estimates of veridicality and internal consistency were obtained for both static and revised subjective probabilities. Subjective revision was again compared to ideal revision but estimates of $p(D)$, the probability of the occurrence of the datum of evidence, as well as $p(H)$, the probability that the hypothesis was true, were required so that internal consistency could be assessed. While measures of veridicality showed the usual conservatism, differences between subjects were three times greater for this measure than for internal consistency. Individuals appeared to exhibit restraints so that deviations from veridicality for the one subjective

probability produced compensatory deviations in the same direction for the other probability. The necessity of such restraint was not mentioned to individuals beforehand. Therefore, as in this present study, there was a general departure from veridicality in probability revision in the form of conservatism, but large differences between individuals in the degree to which subjective probabilities differed from ideal probabilities. Of particular interest was the finding that subjective probabilities were internally consistent, so that when individuals were above or below average in degree of conservatism for one measure it was reflected in values of a related measure.

The question also arises as to whether there is between-task consistency. The consistency mentioned above was confined to a single probabilistic network. It is possible, however, that degree of conservatism in probability revision, or the efficiency of any cumulative process that might be involved, could be the reflection of some underlying general "trait". If this were the case, a ranking of individual performance on one task should correlate highly with the rank-order obtained in a subsequent related task. In particular, a possibility which suggests itself is that degree of Extraversion might be related to the extent that any underlying mechanism exerts an influence on the level of subjective probability.

B. EXPERIMENT TWO : CONSISTENCY OF INDIVIDUAL DIFFERENCES
IN THE ADAPTATION TO A DECREASING SIGNAL PROBABILITY

The preceding experiment provided evidence that a type of cumulative mechanism might be responsible for the degree to which individuals revise their response ratios as the result of a change in external stimulus ratios. The possibility emerged that the extent to which this postulated mechanism exerts an effect is related to the Introversion - Extraversion personality dimension. Degree of Extraversion correlated significantly with degree of response probability revision, as measured by the relation between slope score and the E.P.I. score, and was thus associated with levels of response probability more closely resembling the actual stimulus probabilities.

A second experiment was designed to test whether there was any consistency between tasks in the degree to which individuals adapt to probability change. In addition, it seemed desirable to see whether the obtained correlation between degree of Extraversion and accuracy of response probability revision would persist in a subsequent task.

1. Method

(a) Observers

In reply to letters requesting further participation in the experiment, 15 of the original 26 observers returned for the second study. They were again paid at the rate of \$2.75 per hour for their participation.

(b) Task design and apparatus

As in Experiment 1 observers had to judge on each trial which was the longer of the two vertical lines of the stimulus shown in Figure 5. The stimuli were generated, and responses recorded, on the same equipment as

in the previous study. The confidence keys were not used.

At the beginning of the session there was one practice block of 25 trials in which relative stimulus probabilities were .50. As in the previous experiment the main task then consisted of six test blocks of 100 trials each. There were no breaks between test blocks so that the main task appeared as a continuous sequence of events. Again, as in the previous experiment, the probability of the signal stimulus during each consecutive trial block was .50, .41, .32, .23, .14 and .05, respectively. However, in order to permit an analysis of the rank-order of performance under a single set of circumstances, all observers in this second experiment were tested with the same signal stimulus, namely, the figure with the longer line on the left. It was hoped that this would allow for a stricter test of consistency since half the observers would be performing a task which was quite dissimilar from the original one in which the complementary stimulus was reduced in frequency. Unfortunately, the assumption that all the observers from the original study would return for the follow-up study turned out to be overly-optimistic. As it happened, nine of the observers originally had the longer line to the right as the signal while six had the longer line to the left.

A further difference from the first experiment is that the present task was self-paced. If a response occurred within 2.5 seconds of stimulus onset, the stimulus was immediately terminated and the next stimulus appeared 0.5 seconds later. However, if no response occurred within this period the stimulus was automatically terminated after 2.5 seconds, although responses to this stimulus could still be recorded during the 0.5 second interval which followed. Maximum trial length was therefore 3 seconds. The maximum length of a session, including an interval of about three minutes between the practice trials and the main session (during which a print-out of performance

was checked) was thus approximately 35 minutes.

(c) Procedure

Observers were tested individually and in complete isolation. The same basic instructions were given as in Experiment 1 with the exception that instructions about confidence were omitted. That is, observers were told to respond to the longer of the two lines as quickly as possible, without making too many mistakes, by depressing the right button if the right line was judged longer and the left if the left was judged longer. All responses were to be made with the index finger of the dominant hand. In order to avoid errors of anticipation, the nature of the self-paced task was explained and it was stressed that the index finger should always be returned to its wedge-shaped resting position after each decision response. Following practice trials observers were told that the relative frequency of the two stimuli might, at times, change during the course of the main session but they were given no information as to the nature of this change. Following the main session observers were asked to fill in the Eysenck Personality Inventory.

2. Results

(a) Response probability

(i) General trends. Changes during the session in the probability of correct and incorrect signal responses are shown in Figure 11. As can be seen from the figure, a striking feature of the curve for the probability of making a correct signal response is the drop from the fifth to the last trial block, which contrasts with the apparent steady increase observed during the previous trial blocks. Somewhat surprisingly, however, the change over the first five blocks of trials was not significant ($F = .646$, d.f. = 4/56,

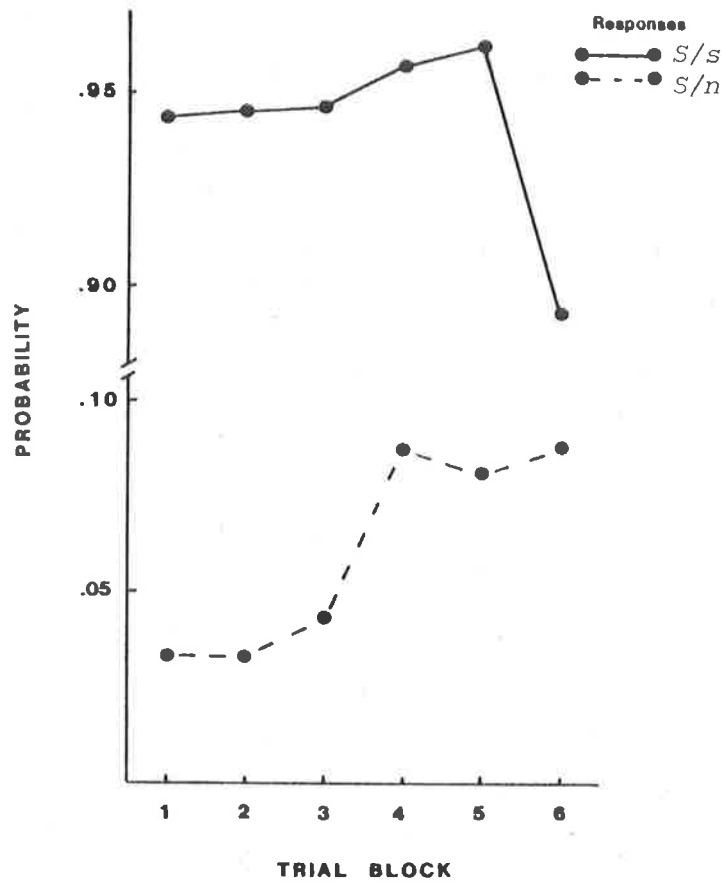


FIGURE 11. Changes in the mean conditional probability of a correct (S/s) and incorrect (S/n) signal response over the six trial blocks.

$p > .6$), nor was that from the fifth to the sixth block of trials ($F = 2.229$, $d.f. = 1/14$, $p > .1$). The latter finding may be due to more erratic responding later in the session (reflected by an increase in the group variance of the probability of making a correct signal report from .004 in the fifth trial block up to .045 in the final block ($F = 12.67$, $d.f. = 1/14$, $p < .001$)).

As in Experiment 1 there appears to have been a strong increase in the tendency to make false reports of a signal during the session, and an analysis of variance, performed over the six experimental blocks, on the probability of making an incorrect signal report found this increase to be highly significant ($F = 5.648$, $d.f. = 5/70$, $p < .0005$). Although there was no

significant change in the proportion of correct signal responses, there was a decrease over the six experimental blocks (corresponding to the increase in false report rate) in the probability of making a correct decision, whether this was a signal or nonsignal response ($F = 3.775$, $d.f. = 5/70$, $p < .005$).

(ii) Changes in response probabilities as a function of local and cumulatively-based signal probability. Figure 12 shows unconditional signal response probability over the six trial blocks. As can be seen from the figure, the unconditional probability of a signal response fell at a greater rate in this second experiment. While the curve for this second experiment still

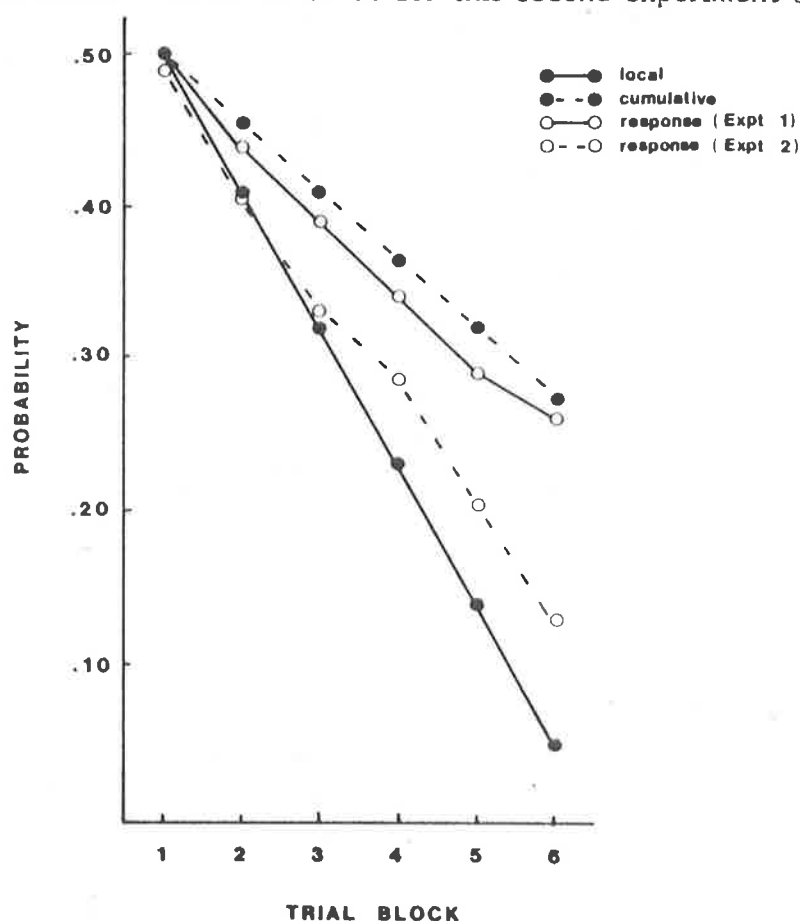


FIGURE 12. Changes in the mean unconditional probability of making a signal response in Experiments 1 and 2 compared with the local and cumulatively-based probability of a signal stimulus over the six trial blocks.

lies somewhere between the cumulatively-based probability of the signal stimulus and its local probability during each block, it lies much closer to the latter. Meanwhile, the unconditional probability of a signal response plotted as a function of local probability was well fitted by a straight line of slope 0.78 and intercept of 0.09. On the other hand, when plotted against cumulatively-based probability both the value of the slope (1.55) and intercept (-0.29) were further from the ideal parameters of 1.0 and 0.0.

As in Experiment 1, a crude goodness-of-fit test was performed, in this case to see whether unconditional signal response probability differed from both local and cumulatively-based signal probability, that is, whether it fell somewhere between the two as suggested in Figure 12. For each observer of the present study, the difference between the obtained slope and intercept and their respective ideal values of 1.0 and 0.0 were computed, in both the case of signal response probability plotted against local signal probability and of signal response probability plotted against cumulatively-based signal probability. As was the case in the previous experiment, when signal response probability was plotted against local signal probability, there was a significant departure from the theoretically ideal values for both slope ($T = 1$, $Z = -3.35$, $p < .001$ (2-tailed)) and intercept scores ($T = 1.5$, $Z = -3.32$, $p < .001$ (2-tailed)). However, in this study, signal response probability plotted against cumulatively-based signal probability also had a function with a slope that differed from 1.0 ($T = 8$, $Z = -2.95$, $p < .005$ (2-tailed)) and an intercept different from 0.0 ($T = 4$, $Z = -3.18$, $p < .005$ (2-tailed)).

The above results, at first glance, seem to indicate that the tendency to over-respond to the signal stimulus may not have emerged in this second experiment. However, it appears from Figure 12 that the tendency to over-respond to the signal became stronger as the session progressed. Since the

signal used was always the stimulus with the longest line to the left, and this was previously found to have a marked bias against it, the initial probability of making a signal response was less than 0.5. This probability then appeared to decline along with local probability at least over the first half of the session. However, over the last half of the session, the response curve seems to become dissociated from the curve for objective local probability in each block.

As initial response probability differed from objective, two difference scores were computed, being firstly the difference between signal response probability and objective local signal probability and, secondly, the difference between signal response probability and cumulatively-based signal probability. The means of these scores for the first trial block, computed from the data for all observers, was compared with their respective means from each successive trial block. As expected, the mean difference between response and cumulatively-based signal probability increased by the second block of trials ($t = 3.80$ (separate variance estimate), d.f. = 20.4, $p = .0005$). However, the mean difference between response and local signal probability had also increased significantly by the third trial block, where the response probability curve first crossed over the local probability curve ($t = -1.95$ (pooled variance estimate), d.f. = 28, $p < .05$). Thus the tendency to over-respond as signal probability declined again increased with time on task, but to a lesser extent than in the first experiment.

(b) Individual differences in the revision of response ratios

The slope of the best-fitting straight line relating unconditional signal response probability to cumulatively-based signal probability across the six trial blocks were calculated separately for each observer. Mean slope was

1.56 as compared to a mean of 0.99 for the same sample of individuals in the previous experiment. A Pearson correlation coefficient of the relationship of the scores from the two studies was not close to significance ($r = -.0307$, $p = .457$). This lack of relationship between slope scores would seem to imply that an individual's performance on the first task was not an indicator of performance on the subsequent task. However, correlations obtained between the slope and Extraversion scores from the two studies suggest that the sample of observers in this follow-up study was not representative of the original experimental group. The correlation of individual Extraversion scores from the two studies was far from significant ($r = .050$, $p = .446$). Among the ten observers who had complete personality data for the first study the mean Extraversion score had increased from an initial value of 9.9 to 12.9, indicating that they had become more Extraverted. This conclusion is further supported by the fact that, unlike that for the entire group, the correlation between personality and performance measures obtained in the original study was not significant for these individuals ($r = .298$, $p = .2$). Such a discrepancy clearly suggests that sampling error may have been operative. At the same time, the correlation between Extraversion and slope score among all the observers in the second study was .414, which has a borderline significance of .062. If the two groups of subjects are accepted as independent, then the correlations between personality and performance scores obtained in the two studies would be highly significant. In any case, there seems to be some grounds for believing that degree of Extraversion may be related to the extent to which individuals are prepared to revise their relative response ratios. While changes in both personality and the rate of response probability revision may occur in the course of a year, there is some evidence for further exploring the possibility of a stable relationship between them.

(c) Signal detection analysis

Changes during the session in d' and β , as computed from the average conditional signal response probabilities during each trial block, are shown in Figure 13. As with the probability of correctly reporting a signal, the trends in these measures parallel the results of the first experiment for the first five trial blocks. There then appears to be a sharp fall in d' and rise in β in the final trial block. Due to the high incidence of correct signal reports, individual values of d' and β could not be computed for each block for

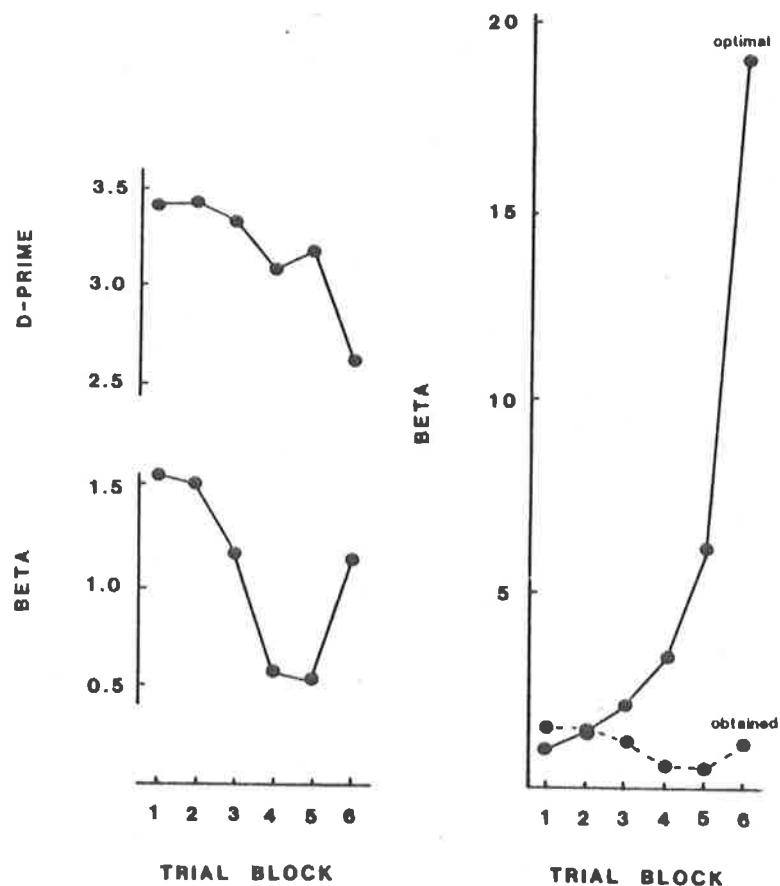


FIGURE 13. Changes in group values of d' and β over the six trial blocks. On the right the values of β that were observed are compared with the optimal values specified by signal detection theory.

statistical analysis. The only possible analysis was based on values of d' and β calculated from data for each half of the session, which largely masked the trends evident in Figure 13. Computed in this way, d' fell from 3.48 to 2.96 as β increased from 1.43 to 2.11. An analysis of variance, which was possible only for the data of seven observers (less than half of the total sample), found the decrease in d' to be significant ($F = 12.685$, $d.f. = 1/6$, $p > .05$), and the increase in β to be nonsignificant ($F = .381$, $d.f. = 1/6$, $p > .3$). As signal detection measures were based solely on the data of the poorer performers, no conclusions can be drawn as to whether the decline in sensitivity was genuine, or whether β was stable. Figure 13 certainly argues for a decline in β , accompanying the decline in signal probability up until the final trial block, although this could not be tested statistically.

(d) Response latencies

(i) Within-session trends. Changes during the session in the latencies of the four types of decision are shown in Figure 14. Latencies for three of the decision categories appeared to increase during the session, while incorrect nonsignal responses appeared to become faster toward the end of the session. Analyses of variance performed on the data across all six trial blocks showed the increase in correct signal and correct nonsignal response latencies to be significant ($F = 4.860$, $d.f. = 5/70$, $p < .001$, and $F = 7.670$, $d.f. = 5/70$, $p < .00001$, respectively). Due to the relative infrequency of false reports and incorrect nonsignal responses, the latencies for these responses were averaged over each half of the session. Analyses of variance indicated no change in the latencies of either false reports ($F = .198$, $d.f. = 1/12$, $p > .6$) or incorrect nonsignal responses ($F = .071$, $d.f. = 1/7$, $p > .7$), though, in the latter instance, this analysis could be based on the data of only eight observers.

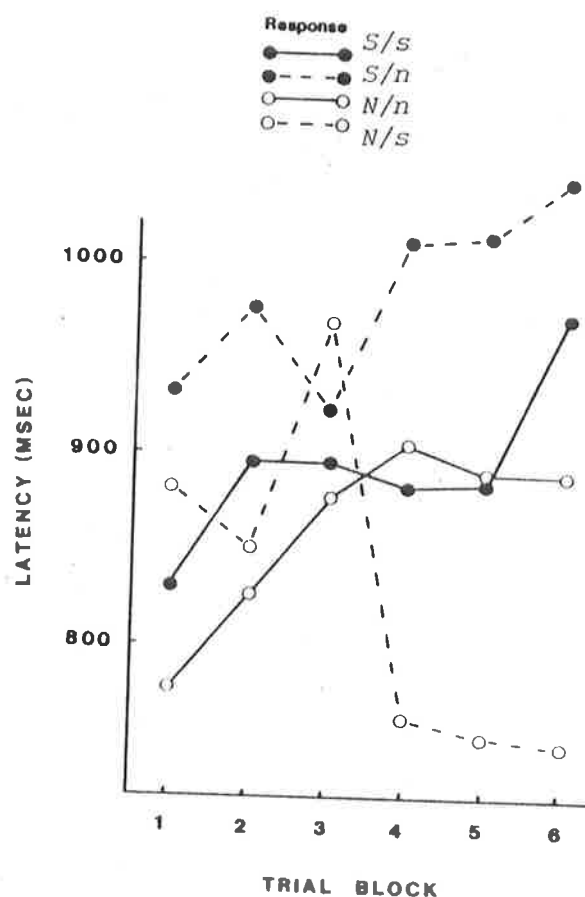


FIGURE 14. Changes over the six trial blocks in the latencies of correct signal (S/s), correct nonsignal (N/n), incorrect signal (S/n) and incorrect nonsignal (N/s) responses.

Although statistically nonsignificant, the apparent decrease in the latency of incorrect nonsignal responses might help to explain the abrupt changes from the fifth to the last test block in the probability of making a correct signal report and in the signal detection measures. In the first test block, mean response latency when signals were incorrectly reported as nonsignals was 882 msec. with a standard deviation of 256 msec. Thus, one standard deviation below the mean, or 68 per cent of the faster responses, were latencies of 626 to 882 msec., not unreasonably fast reaction times. However, the mean and standard deviation for the final block were

747 and 373 msec., respectively. That is 68 per cent of the faster responses were then occurring within 374 to 747 msec., of stimulus onset, and 32 per cent below 374 msec. The only reasonable explanation for these overly-fast responses is that errors of anticipation were occurring. It appears that the self-paced task produced response times that were much lower than those observed in Experiment 1. In this situation anticipatory errors are likely to occur, especially with the short interval of 0.5 seconds between a response and the onset of the next stimulus. Thus as signal stimuli became less frequent, and the mean length of uninterrupted runs of nonsignal stimuli became longer, individuals were then more likely to anticipate a nonsignal stimulus on the next trial.

The increase in the variance of correct signal response probability further suggests that if a nonsignal was not anticipated on a signal trial, the signal would be detected correctly. As only five signals occurred in the final trial block, just two errors of anticipation would have caused correct signal response probability to drop from 1.0 to 0.6, lowering the group mean and increasing group variance. The same two errors in the first trial block would have lowered correct signal response probability from 1.0 to 0.96 and would have had negligible effect on the group statistics. The marked effect of such errors could well have caused a decrease in the probability of making a correct signal response, the range of the group values of the conditional signal response probabilities shown in Figure 11, the increase in β , and the decrease in d' from the fifth to the final trial block.

(ii) Differences among the four response categories. Mean response latencies for the four decision categories are given in Table 4. As suggested in Table 4, correct signal responses were made more slowly than correct nonsignal responses ($t = 2.56$, d.f. = 89, $p = .006$) and incorrect signal responses

TABLE 4. Mean latencies for the four categories of response (in milliseconds).

<u>Signal responses</u>		<u>Nonsignal responses</u>	
Correct (S/s)	Incorrect (S/n)	Correct (N/n)	Incorrect (N/s)
893	990	860	847

were made more slowly than incorrect nonsignal responses ($t = 2.63$, d.f. = 36, $p = .007$). These two findings are consistent with the bias found against making "left longer" responses. Figure 13 indicates that, at the beginning of the task, β was somewhat greater than 1.0 (the value which would have resulted had no differential bias existed) and that it did not decline past this equal-odds value until the fourth test block.

As would be expected on a latency function hypothesis, correct signal responses were faster than incorrect signal responses ($t = -4.01$, d.f. = 68, $p < .0005$). However, the contradictory finding that incorrect nonsignal responses were shorter than correct nonsignal responses did not quite reach significance ($t = -1.58$, d.f. = 47, $p = .06$), and could be explained as the result of anticipatory errors which occurred later in the task.

3. Discussion

(a) Consistency in the degree of revision of response ratios

For the present, the question of whether there is between-task consistency in the degree to which individuals revise their response ratios in response to a change in objective stimulus ratios cannot be fully resolved. The degree of revision that was shown in the first study was not related to the degree of revision in the second study, i.e. individual performance on the first task was not related to its rank-order in the second task.

However, it appeared that individuals had also changed with respect to their measured degree of Extraversion. Changes in this personality variable, which was originally shown to be related to task performance, thus appears to have nullified the attempt to determine whether there is consistency in performance between tasks. Moreover, the individuals who returned for the follow-up study were not representative of the original sample since in the first study they did not show the personality-performance relationship that was true for the group as a whole.

There is, however, some evidence to indicate that the extent to which probabilities are revised is the reflection of an underlying personality trait. Degree of Extraversion, as obtained in the second study, showed a marginally significant correlation with the degree of probability revision. Considering the small size of the experimental group, which bore no relationship to the original group, it would appear to be premature to dismiss this finding.

(b) Response to a change in signal probability - a proposed target for the adaptive process

The most striking feature of the present study is that although the observers were probably prepared for a decline in the probability of one of the stimuli, they did not exhibit a tendency to become more biased against making its report. Instead, the trends observed largely parallel those of Experiment 1. As the relative probability of a signal declined, the tendency to over-respond was again displayed and emerged later in the task following a period of somewhat accurate adjustment in response ratios. This was accompanied by a steady increase throughout the task in the probability of making an incorrect signal report, and an increase in the probability of a correct signal report. These trends persisted until the final trial block where

the latter trend was obscured by errors of anticipation.

Although response probability revision was more accurate than that observed among the same individuals in Experiment 1, previous experience with a decline in signal probability did not lead to an increase in the criterion used to report the presence of a signal, as would be demanded by a response optimization model. Rather, a downward trend was again indicated during most of the task. Combined with the results of Experiment 1 this further supports the notion that the initial adjustment in the decision criterion in response to a change in signal probability is in the opposite direction to that specified by the ideal observer hypothesis.

The observed changes in β were in the opposite direction to those usually reported in vigilance research where signal probability is decreased abruptly from an expected level at the beginning of a task and held constant. However, the results of the present experiments suggest that the increase in β observed in such experimental situations (e.g. Colquhoun & Baddeley, 1967; Williges, 1969, 1973) may occur from an initially overly-low value. That is, the initial reaction to a lowered signal probability is the adoption of a less stringent criterion which is then increased in stringency as a result of subsequent experience in the task.

Such an initial change would be imperceptible to the usual type of data analysis employed in vigilance situations, where changes in performance are compared against the initial level of performance exhibited in the task. For example, if this initial level follows a drop in objective, or expected, signal probability, it may represent an adaptation from a preceding level, such as one encountered in practice. A trial-by-trial analysis of performance very early in the task would have to be compared with performance during the initial learning of subjective probabilities, in order to detect whether such an adaptation had taken place.

For the present, there seems to be some justification for the hypothesis that when a change in signal probability is encountered, the observer will adopt a value of the decision criterion that results in a matching of signal response probability with the cumulatively-based probability of signal occurrence.

(c) Response maintenance - a clearly inadequate interpretation of the present results

The above hypothesis can be compared with Williges' (1969) response maintenance hypothesis. This suggested that the tendency for observers to display high values of β at a high level of signal probability and low values at a low signal probability reflected over-responding or under-responding in order to match the probability of a signal response with the signal probability implied by training. According to this notion, if signal probability declines from the level at which observers are trained, there must be an increase in either, or both the conditional probabilities of a signal report in order to maintain the unconditional probability of a signal report near its previous level. The observer can achieve this by one of three strategies, two of which involve a decrease in the measure β , and so agree with the present results, but only one of which agrees with the results obtained with the other response measures.

(i) Strategy one; "guessing" on both signal and nonsignal trials. The first strategy is the least rational for a supposedly intelligent decision-making system. It involves making a "guess" on each trial as to whether a signal was presented, irrespective of what response was favoured by the evidence on that trial. For instance, if signal probability was initially 0.5 and then decreased, the observer will randomly report that 50 per cent of the stimuli that are

presented are signals, so that a constant probability of 0.5 is attached to each event being a signal. In this way unconditional signal response probability will remain at 0.5, but there will only be a 50:50 chance that any stimulus, signal or nonsignal, will be reported correctly. To give a numerical example, let us suppose that initial signal probability is 0.5, that the observer begins by effectively processing sensory information, and that the probabilities of a correct detection and false alarm are 0.9 and 0.1, respectively. If signal probability then declines to 0.1, and the observer resorts to guessing on all trials, correct detection and false alarm probability would then each be 0.5 and this would result in a drop from 0.9 to 0.5 in the probability of a correct decision. While measured β would remain at 1.0, d' would drop from 2.56 to 0.0. Thus, the consequence of this behaviour would be a decline in measured sensitivity. This interpretation is clearly inconsistent with the increase in detection rate and stable d' which were observed in Experiment 1.

(ii) Strategy two (response maintenance): guessing only when the evidence favours a nonsignal response. The second strategy is to attach a fixed probability of making a signal report to each of the stimuli that would otherwise give rise to nonsignal reports. That is, on each trial that the amount of evidence is not great enough to result in a signal report, the observer guesses, with some fixed probability, as to whether a signal has been presented. The consequence of this procedure would be a disproportionately large increase in the probability of false reports relative to a concomitant increase in correct detections.

To illustrate, let us suppose again that signal probability is decreased from 0.5 to 0.1, and that the probability of a correct detection is initially 0.9, and that of a false alarm is 0.1 (corresponding to an initial, and ideal, equal-odds β). Prior to the decrease in signal probability, 45 signals and 5 nonsignals

in a block of 100 stimuli would be reported as signals, so that unconditional signal response probability would be 0.5. If signal detectability and response bias are unaffected, then, at the new signal probability, 9 of the 10 signals and 9 of the 90 nonsignals in a block of 100 stimuli will be reported as signals. Unconditional signal response probability is now 0.18, and, in order to maintain it at its previous level of 0.5, 32 additional stimuli must be reported as signals. Hence, 39 per cent of the 82 stimuli that would otherwise lead to nonsignal reports will have to be reported as signals.

If on each trial that the evidence is not great enough to exceed the criterion for a signal response, the observer makes a guess as to whether the stimulus was in fact a signal, then each of the 82 stimuli will have the same probability, 0.39, of being reported as a signal. Consequently, the sole remaining signal stimulus will have a probability of only 0.39 that it will be detected, while 31.61 (on average) of the remaining 81 nonsignal stimuli will be falsely reported as signals. The probability of a correct detection then increases slightly from 0.90 to 0.939, while that of a false report increases from 0.1 to 0.451. The increase in the unconditional probability of a signal response is thereby achieved mainly through an increase in false alarm rate. The result of such a substantial increase in false alarm rate, with a marginal increase in detection rate, is to reduce both signal detection measures. In the present example β declines from 1.0 to 0.31 and d' from 2.56 to 1.67.

It should be noted that the arbitrary choice of initial conditional response probabilities did not produce a unique result. For example, if the same sort of guessing procedure had been employed with an initial detection rate of 0.8 and false report rate of 0.2, then detection rate would have only increased to 0.865 while false reports rate increased to 0.459. Corresponding to this are decreases from 1.0 to 0.551 in β and from 1.683 to 1.207 in d' .

In the present studies there were no such highly disproportionate increases in false alarm rate (see Figures 6 and 11), nor any decline in measured sensitivity to accompany the decline in β in Experiment 1 where the signal detection analysis of performance was by far the least dubious. Thus the observer was not using any simple guessing procedure to increase the proportion of signal reports. This latter procedure of guessing only on those trials which favour a nonsignal report is the essence of Williges' (1969) response maintenance interpretation, which suggests that β reflects an alteration in response ratios rather than a deliberate adjustment of the decision criterion.

(iii) Strategy three: cumulative probability-matching through a deliberate systematic change in the decision criterion. The third possibility is that the observer increases the unconditional probability of a signal response by accepting less evidence to indicate that a signal is present, but, unlike the guessing process, requires that a specific minimum amount be observed before a signal is reported. Instead of all observations that previously would have led to a nonsignal report now having a fixed probability of leading to a signal report, only those observations which produce a certain amount of sensory evidence will be reported as a signal. There is then an orderly leftward movement in the placement of the decision criterion along the sensory continuum.

Again the increase in the probability of a false report will exceed that of a correct detection, since the decision criterion will pass a greater proportion of the nonsignal distribution than the signal distribution as it moves to the left toward the mean of the nonsignal distribution and over the tail of the signal distribution as shown in Figure 15. However, the rate of increase in the two conditional response probabilities will not be nearly so

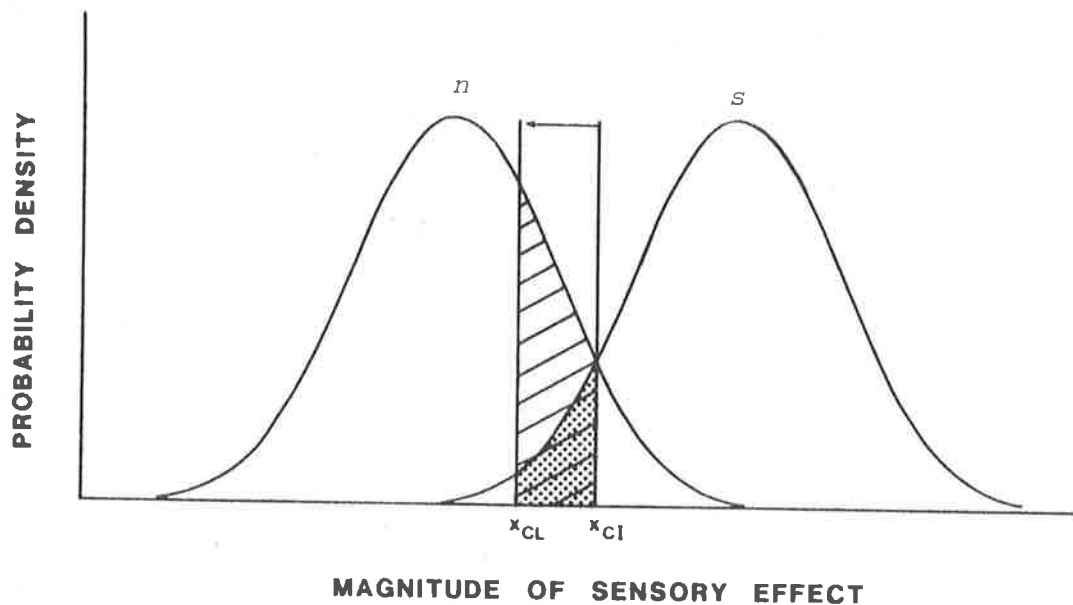


FIGURE 15. Effect of a change from an intermediate to a more lax decision criterion. As the cutoff value moves from x_{CI} to x_{CL} a smaller proportion of s (shaded) is passed than of n (diagonal lines). The increase in the probability of a correct signal response is then less than that of an incorrect signal response.

disparate as in the examples given for the previous two response strategies. Instead, conditional probabilities will more closely resemble the increases observable in Figures 6 and 11, especially if an intermediate criterion is adopted initially, as in these examples. Furthermore, the increase in unconditional signal response probability is achieved through a decrease only in the value of the decision criterion, or β , with sensitivity unaffected, and so agrees with the results found with the signal detection measures.

This strategy closely resembles Craig's (1978) explanation of vigilance decrement in terms of the 'probability-matching' hypothesis as applied in psychophysical experiments. This model supposes that the observed shifts in β reflect a tendency for the observer to adjust his signal : nonsignal response

ratio toward the objective signal : nonsignal stimulus ratio. However, Craig's model would predict that during any vigilance task in which there is no change in signal probability during the event sequence, the observer would adjust his response ratio so that the probability of making a signal decision would approach the probability of signal occurrence during the task itself, irrespective of any earlier experience at a different level of signal probability. The present hypothesis, however, states that when there is a discrepancy between the present local probability of a signal and an earlier level of signal probability, the observer will alter his decision criterion in order to match unconditional signal response probability with the cumulatively-based probability of signal occurrence. If no discrepancy has taken place, the present local probability of a signal will match the cumulatively-based signal probability so that Craig's (1978) model and the present hypothesis are then equivalent. For this reason the two approaches may perhaps be merged into a single 'cumulative probability-matching' hypothesis.

CHAPTER 3

A. 1. FURTHER EVIDENCE OF NON-OPTIMAL REVISION IN β

As discussed in the previous chapter, Williges (1969) suggested that an inverse relationship between actual and optimal values of β might be due to a tendency for observers to maintain an initial proportion of signal responses. In an analogous way the results of Experiments 1 and 2 suggest that the observer may change his value of β in such a way that his current proportion of signal responses more closely matches the cumulatively-based proportion of signals presented through a session. Besides these experiments, several recent studies have confirmed the apparent paradoxical change in β when a change in signal probability is encountered. The first of these (Parducci & Sandusky, 1970) was concerned with the effects of different signal probabilities on overall performance in a situation where the observer could be assumed to have an initial subjective probability of 0.5. Two subsequent studies by Williges (1973) and Embrey (1975) altered signal probability during the experimental session itself and, as in Experiments 1 and 2, measured the effect of a change in current signal probability. All three studies indicate that, following an alteration in signal probability, β is revised in the opposite direction to that predicted by a model in which the goal of the observer is an optimization of β . Before attempting to outline a model which would give an integrative account of these results, it will be convenient to examine the three studies in more detail.

In the first study, Parducci and Sandusky (1970) employed signal probabilities of 0.2, 0.5, and 0.8 in both a recognition and a discrimination task requiring the identification of "loud" or "soft" tones. This task was then free of any effects due to asymmetry of the payoff matrix. Although a bias was found for louder judgements, the same result was

found for both response alternatives in both types of task: as the probability of a stimulus was increased the probability of reporting that stimulus, conditional on either of the two stimulus alternatives, decreased. As discriminability was not affected by signal probability, β can be inferred to have changed in the opposite direction to that which would be predicted on the basis of its optimal value. No analysis was carried out by Parducci and Sandusky to see whether β changed during the session of 300 trials and, if so, in what direction. However, the overall tendency to over-respond to the stimulus which decreased in probability supports the response maintenance interpretation of Williges (1969) as discussed in the final section of Chapter 2. As individuals were unaware of the relative probabilities at which they were to be tested, and the design was symmetric, initial subjective probabilities might be assumed to have been 0.5. If this were the case, it seems likely that initial values of β might then have been too lax for the low probability condition and too stringent for the high probability condition. Although measures of β were not given by Parducci and Sandusky, this interpretation gains some support from the finding that response ratios were approximately 0.5 at all probability levels. However, no conclusion can be drawn as to whether this ratio was maintained throughout a session.

In the second of these studies, which has been discussed in the previous two chapters, Williges (1973) measured the effects on β of increasing signal probability from an initial level of 0.1 to one of 0.5 during a vigilance task in which observers were aware that such a change in probability would occur. If the observer's response strategy were to optimize β then he would be expected to adopt an initial value of 9.0 for β and then to decrease this toward 1.0 following the decrease in signal probability.

However, β remained almost invariant throughout, starting at a value somewhat less than 2.0 and finishing at a value slightly greater than 2.0. Even though observers had been trained at a signal probability of 0.1, the initial group value of β was closer to the optimal value which would be indicated by the signal probability they were told would occur "at the end of the session". Although the initial value of β was unexpectedly low, there was no further tendency for β to approach its ideal value of 1.0 when signal probability did in fact decrease. This maintenance of β can be seen as an attempt to maintain a reduced probability of making a signal response in the face of a higher objective signal probability. Indeed, the slight nonsignificant movement upwards in β was accompanied by a significant decline in detection probability. Thus, although detection rate at the beginning of the session was virtually the same as that of a group who monitored under a constant signal probability of 0.5, this value had dropped by the end of the session to a point midway between this and the value for another group exposed to a constant signal probability of 0.1.

In the third study, Embrey (1975) provides evidence that with adequate instructions and prior knowledge of variations in signal probability, β will change in the direction specified by the ideal observer model; however, without such cueing the tendency is for β to change in the opposite direction. Observers highly trained at a signal probability of 0.5 were tested at probabilities of 0.1, 0.5, and one which changed from 0.5 to 0.2 during the session. As would be expected, the values of β obtained were close to the ideal when tested at the 0.5 probability. However, despite prior knowledge that signal probability would be low and feedback after every 100 trials, when observers encountered the low signal probability of 0.1, the group β of 1.82 was only marginally above the ideal value of 1.0 adopted during training, and much

less than the optimum of 9.0. Although values of β obtained at more extreme signal probabilities might be expected to be conservative as compared to optimal values (Green & Swets, 1966, Chapter 4), within-session change indicated a tendency at first towards, and then away from optimality. Specifically, an initial rise to 3.14 followed by a decline to 1.82 was recorded among those observers who monitored at the constant signal probability of 0.1. The signal : nonsignal response ratio which would have accompanied such a change in β then, at first, dropped toward the level specified by the objective stimulus ratio and, subsequently, increased. Thus, neither were response ratios maintained, nor was there a tendency for β to approach its optimal value. A similar result emerged in the changing probability condition when no warning was given that signal probability would change. During the first 200 trials β , at 1.18, was close to the value specified by the testing and training signal probability of 0.5. When this probability dropped to 0.2 over the last 300 trials, β declined to 0.80 rather than increased toward its optimal value of 4.0. Such an effect was not found, however, when observers had prior knowledge that the change in signal probability would occur.

Like the experiment of Williges (1969), and those reported in Experiments 1 and 2, these three studies all give evidence that β might be revised in a direction opposite to that specified by the ideal observer hypothesis. In order to account for the non-optimal behaviour of β when a change in signal probability is encountered, and the tendency to maintain response ratios in certain situations of objective stimulus probability change while adjusting them in others, a model was formulated in conjunction with D. Vickers. This model is presented with the results of Experiment 1 by Vickers, Leary and Barnes (1977), and represents the joint product of Vickers' interest in adaptive decision mechanisms and the writer's estimation of the

importance of cumulatively-based signal probability. The version presented by Vickers, Leary and Barnes is considered by them in the context of alternative forms of psychophysical decision mechanisms to that envisioned by signal detection theory (Vickers, 1978; Vickers, 1979). For present purposes, however, it is perhaps more convenient to outline a version which is compatible with the theoretical framework of signal detection theory.

2. AN ALTERNATIVE ADAPTIVE RESPONSE STRATEGY - MINIMIZE THE DISCREPANCY BETWEEN LOCAL AND CUMULATIVELY-BASED SIGNAL PROBABILITY (RESPONSE STABILIZATION)

The model is based empirically on the different trends in β which result from an abrupt change in signal probability, on the one hand, and those which accompany a gradual change, on the other. The particular type of manipulation will determine the pattern of the relationship between cumulatively-based signal probability, C (the probability of the signal during the entire preceding event sequence) and its local probability, L (the actual probability for the immediately preceding sequence of events). The results of Experiments 1 and 2, in which signal probability decreased steadily from an initial "expected" level, rather than decreasing abruptly at the beginning of the task and then remaining constant, differed from those earlier studies in which the step reduction in signal probability led to an increase in β (e.g. Colquhoun & Baddeley, 1967; Williges, 1969). As shown below in Figure 17, both manipulations lead to a situation in which local signal probability, L , falls below the cumulatively-based probability, C , of a signal. As a result the value of $(L-C)$ is negative. However, the absolute magnitude of this difference, $|L-C|$, increases during a session in the former case (a ramp reduction in signal probability), but decreases in the latter (a step reduction in signal probability).

Since either of the two response alternatives of Experiments 1 and 2 could be arbitrarily regarded as a "signal" for the purposes of analysis, the observed effects of a gradual increase in signal probability are also the opposite to those hitherto observed when signal probability is abruptly increased. Both procedures lead to a situation in which $(L-C)$ takes on a positive value. The magnitude of $|L-C|$, however, again increases during a session in Experiments 1 and 2, while it decreases following the step increase in the experiment of Williges (1969). As suggested by Vickers *et al.* (1977, p. 699), these opposing changes in the degree of signal response bias appear to depend on whether the discrepancy between the local and the cumulatively-based probability of a signal response is positive or negative, and whether this discrepancy is increasing or decreasing. In order to explain these effects, these authors suggested that an observer may be attempting to "minimize any discrepancy between the local and the cumulative probability of making a signal response" (Vickers *et al.*, 1977, p. 699). The sequence of operations for an adaptive process which will achieve this goal is shown in Figure 16. According to this model, at any point in an experimental session the entire preceding response sequence, including demonstration and practice, as well as experimental trials, is used to provide the information for a subjective estimate of C , while a similar estimate of L is calculated over the (usually shorter) sequence of the last n trials. According to this version, adaptation, in terms of the adjustment in degree of response bias or β , should occur whenever $|L-C|$ exceeds some minimum critical level, M . As foreshadowed above, the direction of the adjustments depend upon whether $(L-C)$ is positive or negative, and whether the current amount of discrepancy, $|L-C|_j$, is greater or less than the amount last noted by the system, $|L-C|_i$.

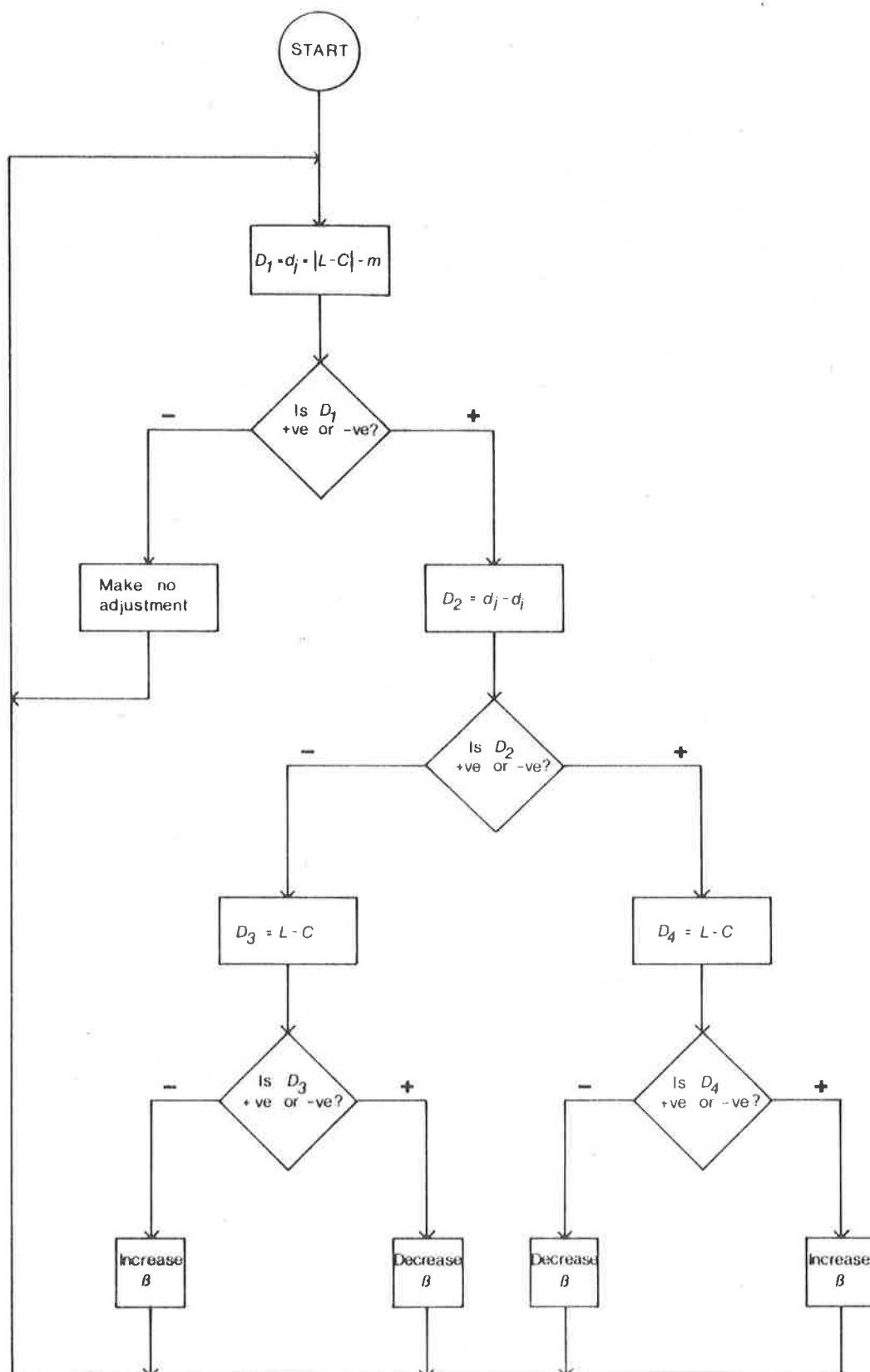


FIGURE 16. Flow diagram of the proposed response stabilization process for the minimization of discrepancies between the local (L) and cumulatively-based (C) probability of making a signal response. The minimum tolerable discrepancy is given by m . (From figure 8 of Vickers, Leary and Barnes (1977).)

When a discrepancy between L and C is first encountered by the system, the difference between this and any previous discrepancy will always be positive so that the decision rule is that the decision criterion should be increased if L exceeds C and decreased if C exceeds L . As illustrated in Figure 17, the values of L and C converge following a step change in signal probability so that there will be an adjustment in β in the opposite direction upon noting the next discrepancy. Traditional vigilance decrements, such as those obtained when observers begin the session with a "falsely-high"

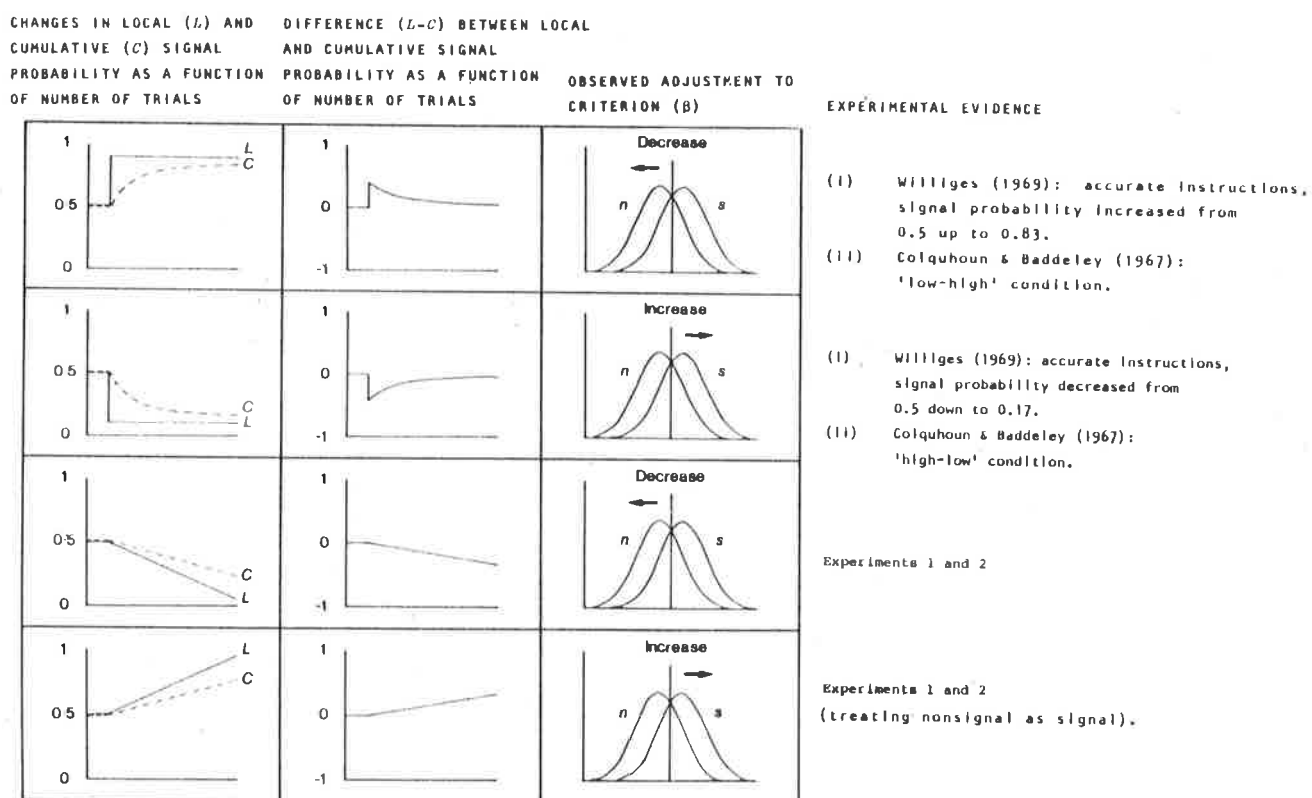


FIGURE 17. The general form of changes in the objective local (L) and cumulative (C) probability of a signal from training to testing, variations in the difference ($L-C$), and adjustments to β , as inferred from the experimental evidence cited in the right-hand column. (From figure 7 of Vickers *et al.*, 1977.)

expectancy ($C > L$), are thus seen as the more gradual recovery of β following a rapid decrease when the discrepancy is first noted. Conversely, the tendency for β to decrease during a session when observers are tested at a higher probability than that encountered during training ($L > C$) would be seen as a downward recovery after a rapid initial increase. This hypothesis is therefore consistent with the notion of "rapid adaptation" as put forward in the previous chapter. In the case of the ramp change in signal probability encountered in Experiments 1 and 2, the values of L and C continue to diverge so that adjustments in β continue to be in the same direction.

The adjustments specified by the model may also be interpreted in terms of an expectancy-type approach. That is, the observer "expects" to make a certain proportion of signal responses. If the present proportion seems too low, so that it is felt that some signals are being missed, the initial reaction is to be more willing to make a signal response and so increase the proportion. On the other hand, if this proportion suddenly increases, so that the observer feels he is reporting too many, the initial reaction is to be less willing to make a signal response and thereby decrease the proportion. These initial 'gambler's fallacies' will be maintained until either (1) the difference between the proportion of reports being made and the expected (cumulatively-based) proportion begins to converge, at which time response bias will change in the opposite direction, or (2) there is a coincidence of local and cumulatively-based signal probability, and no further criterion adjustments are necessary.

This model is clearly based on the assumption of two distinct probability stores. Besides the evidence for a long-term cumulative store of probability information discussed previously, there is some indication

that recent events or responses are also represented in a distinct store. For example, Howell (1970) required individuals to estimate the proportion of their responses which had been made correctly, as well as to estimate the relative stimulus proportions. He found that a record of accuracy was kept which was biased toward the most recent events. This occurred in conjunction with a cumulative store of relative stimulus frequencies for the entire sequence. In another study also consistent with these assumptions, Beach, Rose, Sayeki, Wise and Carter (1970) employed a task in which individuals were required to estimate the relative proportion of two events as well as predict which event would occur next. Estimates approached the actual proportion more rapidly than the proportion of predictions and it was proposed that, whereas estimates reflect the learning of actual proportions, predictions are based on cues such as the preceding run of events.

Both these studies suggest that there is a long- and short-term store of relative frequency information. In support of this, Howell (1973) argued that two frequency processes were involved in the storage of frequency information about nonverbal stimuli: "a rapid, short-term grouping (**"subitizing"**) of like events, and a longer term build up of strength or multiple traces" (p. 49). A relationship between two such probability stores was advanced in a model developed by Ward Edwards and Gordon Robinson (reported in Robinson, 1964). It was proposed that in the tracking of a time-varying probability individuals maintain a short running average of the immediately preceding events which is compared with the current probability estimate. Adjustments are then made so that the new probability estimate takes on a value intermediate between the previous estimate and the running average. The discrepancy between the previous estimate and the running average is thereby reduced, as is the goal of the observer in the present model.

In terms of the present model, the process can be termed one of response stabilization, since it serves to reduce any discrepancy between the present and long-term proportion of responses, and so brings about a state of equilibrium between the two.

The present approach might also be used to explain the relationship between degree of Extraversion and the extent to which signal response probability was revised, in terms of either of three factors. Firstly, the adjustments in the decision criterion in response to a perceived discrepancy could be smaller among Extraverts. Secondly, Extraverts might adopt a higher value of M so that adjustments would be made on the basis of larger discrepancies. Lastly, the estimate of L by Extraverts might be based on a larger number of trials than for Introverts so that it would lie closer to the value of C and give rise to smaller discrepancies. Any of these suppositions could also be used to explain why response probabilities declined more rapidly in Experiment 2. Previous experience with the task could have led to smaller criterial adjustments when a discrepancy was noted, or to a higher limit of tolerable discrepancy. Perhaps an even more plausible explanation is that the event rate in the second study, which was higher than that of the first by an average factor of at least four, may have led to an estimate of local probability that was based on a larger number of trials. This could explain the finding that the tendency to over-respond emerged to an appreciable extent only later in the task. At the beginning of the task both local and cumulatively-based probability might conceivably have been based on a similar number of trials so that the sequence over which L was calculated came to differ substantially from the total sequence only later in the task. Any discrepancies between them would be expected to be smaller and might be noticed later in the response sequence than in Experiment 1.

3. EXPLANATION OF THE NON-OPTIMALITY OF β IN OTHER STUDIES

The response stabilization model can also be used to analyze the previous studies in which a change in signal probability led to revision of β in the opposite direction to that specified by a response optimization process.

For example, if the reasonable assumption is made that observers in the Parducci and Sandusky (1970) experiment began the task with a subjective probability of 0.5, then the present hypothesis would lead us to expect that the initial adjustment of the observers in the low probability condition would be the adoption of a lax criterion, while the converse adjustment would be expected under the high probability condition. Furthermore, since in the absence of feedback, any revision in the estimates of L and C is assumed to be based on response rather than stimulus proportions, it is possible that there may have been little need for adaptation later in the task. The high error rates of the task, a consequence of initial adaptation, then may have been maintained throughout. Such an explanation would also fit Williges' (1969) results where observers adopted values of β seemingly to respond at those rates established during training.

Turning to the results of Williges (1973), these may also be explained on the present hypothesis if the added assumption is made that following instructions which stated that signal probability would be 0.5 by the end of a session, observers began the task with a value of β appropriate to a signal probability intermediate between 0.5 and the training probability of 0.1. As in the case of training at an overly-high probability, an initial decrease in β followed by a tendency to increase would be expected. The empirically-found increase in the number of signal responses that accompanied the increase in signal probability midway through the session

may have caused L and C to coincide so that no further criterion adjustments would be initiated. Hence the result was an inappropriately low β which then showed a slight tendency to move upwards.

An initial drop in the signal decision criterion followed by an increase would also be predicted for Embrey's (1975) low probability condition. Although observers had been told to expect a low incidence of signals, there had been 3000 training trials at a signal probability of 0.5 so that the initial expected probability, or C , was considerably greater than the objective level of 0.1. A further, previously unexplored, possibility is that β may be determined by cumulatively-based signal probability. If so, after the first 100 experimental trials observers would have encountered 1510 signals in the entire preceding sequence of 3100 trials, corresponding to a cumulatively-based signal probability of 0.487. An ideal observer hypothesis based on this probability, rather than the local probability of 0.1, would require a β of 1.05. This is very close to the one of 1.16 which was found empirically, and far from the one of 9.0 that is required by the ideal observer hypothesis based on local signal probability. It then seems reasonable that β may be based on a subjective estimate of cumulatively-based signal probability calculated from a "moving window" of, say, at least the preceding two to three hundred trials. When the estimate provided by this window approaches the local probability of a signal, then the value of β will tend to approach the value which would be based on local signal probability alone. Hence, in Embrey's (1975) low probability condition, β rose to 3.14 over the second 100 trials, before a further decline which may have been a function of the feedback provided in the task (unfortunately, a "no feedback" condition was not reported). The initial overly-low β , followed by an increase, could then have been the reflection of either an attempt by the proposed response stabilization process to minimize the discrepancy between the signal probability

encountered at the beginning of the task and the cumulatively-based probability of signal occurrence, or it might simply be evidence that observers adopt values of β appropriate to an estimate of signal probability based on a larger number of trials than that over which the present local signal probability is calculated by an experimenter.

The strongest evidence for response stabilization in Embrey's (1975) study is contained in the changing probability condition, where no warning was given of the change although feedback was still provided. As signal probability was initially at the same level as that expected through training and instructions, 0.5, and was then decreased abruptly after the second 100-trial block, a drop in β when the new signal probability was encountered, followed by a subsequent increase, would be expected. In full agreement the actual results were values of β of 1.31 and 1.04 for the first two trial blocks, and a drop to 0.59 in the third test block followed by an increase to 1.13 and 0.67, respectively, in the final two blocks.

B. EXPERIMENT THREE : PERFORMANCE CHANGE FOLLOWING A PERIOD OF TIME-VARYING SIGNAL PROBABILITY

The uncertain effects of instructions and knowledge of results discussed in the latter two experiments reviewed in the previous section make any firm conclusion difficult. Therefore the present experiment was undertaken to investigate response to change in signal probability, both initially and during a task, in a situation which was relatively free from biasing factors on the decision criterion, and was unlikely to give rise to any response strategies on the part of observers which could make interpretation difficult. The main objective was to produce alterations in signal probability that would lead to contrasting predictions by the ideal observer

hypothesis (response optimization), the response maintenance approach of Williges (1969), and the response stabilization model outlined above. Such a situation can occur, for example, when signal probability is decreased gradually in a ramp fashion and then held constant. Since the shifts in the decision criterion which are specified by the response stabilization model can be inferred directly from a plot of $(L-C)$, the changes in criterion can be predicted in a straightforward fashion, and are shown in Figure 18.

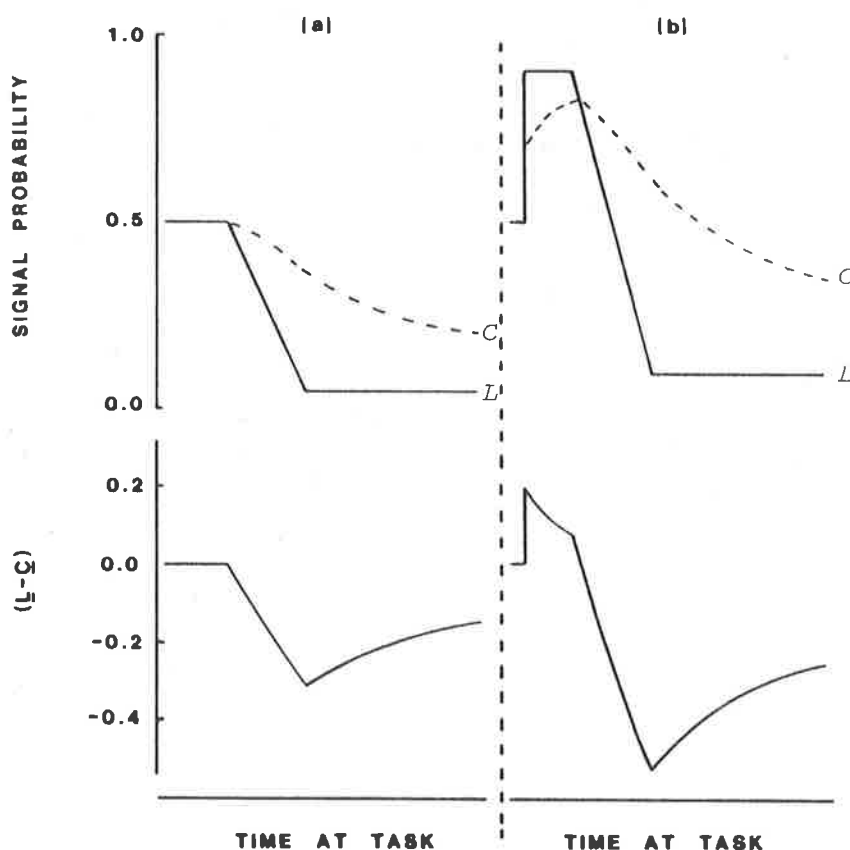


FIGURE 18. Changes in local (L) and cumulatively-based (C) probability of signal occurrence, and the difference ($L-C$), when signal probability is reduced in a ramp fashion and then held constant. The changes in bias specified by the response stabilization model directly follow the curve of $(L-C)$.

(a) Signal probability is reduced from the training level of .5.

(b) Signal probability is increased above the training level of .5 and is then reduced.

Two different situations are depicted in Figure 18. In the first signal probability is maintained for some time at the training level of 0.5 after which it declines to a lower level at which it then remains stable. None of the approaches so far discussed would specify any change during the initial phase of constant probability. During the period in which signal probability declines the ideal observer hypothesis would predict an increase in the value of the decision criterion while both the response stabilization model and the response maintenance hypothesis would predict the reverse, as shown in Figure 19. Over the period of constant probability following the decline, the ideal observer hypothesis and response maintenance would both specify that no

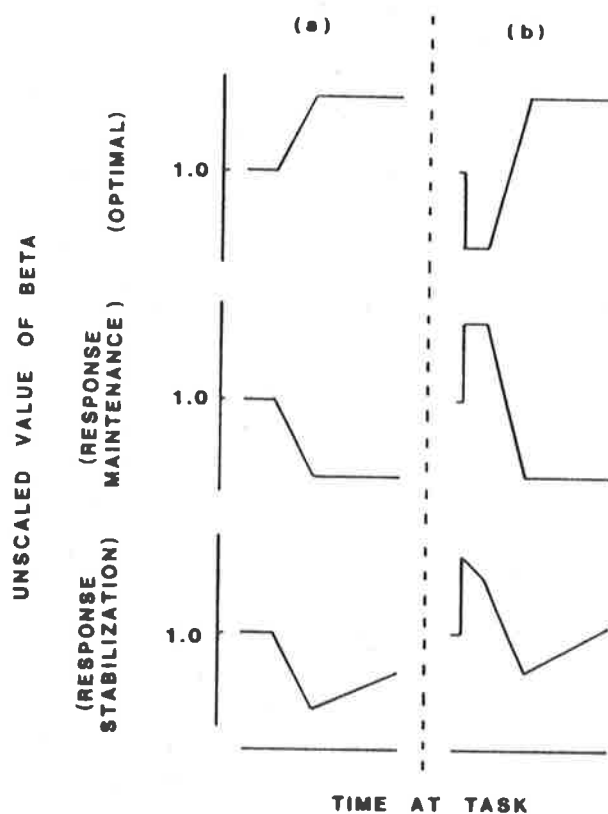


FIGURE 19. Changes in β predicted by each of three theoretical models for the two situations of time-varying signal probability depicted in figure 18. Values of β are unscaled around 1.0.

change in the decision criterion should take place. On the other hand, the response stabilization model would predict an increase in the value of the decision criterion, since the values of L and C will begin to converge and $(L-C)$ thereby increases. In Figure 18 the values of L and C are plotted on the basis of objective signal probability. The response stabilization model supposes that these are based on the actual proportion of responses made. However, where performance, as in the present experiments, is highly accurate, the pattern depicted in Figure 18 can be taken as a close approximation to what would be expected on the basis of actual response probabilities.

In the second situation depicted in Figure 18, signal probability is initially increased from its training level of 0.5. Here the ideal observer hypothesis would predict a decrease in the decision criterion, while, as shown in Figure 19, both the response stabilization model and the response maintenance hypothesis would predict an increase. After the initial increase in signal probability, it is again decreased in a ramp fashion and then held constant. The trends in the value of the decision criterion predicted by each of the models will then correspond to those made for the first situation (see Figure 19). The ideal observer hypothesis and the response stabilization model thus differ in their predictions for each stage of probability change while the crucial divergence between the response stabilization and the response maintenance approach occurs in the final period of constant signal probability.

1. Method

(a) Task and stimuli

A discrimination task that avoided the use of stimulus definitions such as "signal" or "critical event" was again chosen in order to eliminate possible effects of response bias. On each trial observers were required

to judge which was the longer of the two vertical lines of a π -figure stimulus, similar to that used in Experiments 1 and 2 (Figure 5), presented on the screen of a television monitor. The longer line was always 55.5 mm and the shorter was 54 mm. The lines were 20 mm apart and viewed at a distance of 4.4 metres so that the differences in line lengths subtended a visual angle of .02 degrees. After the primary decision response a rating of confidence in its accuracy was made before the next trial began. Each response alternative, "right-longer" or "left longer", had its own 5-category scale of confidence.

(b) Apparatus

The stimuli were generated via a graphics program contained in a D.E.C. PDP8/L computer. The side on which the longer line would appear on any trial was determined by a disk-based D.E.C. PDP8/E system. This second computer determined the amount of time that a stimulus would be visible (2 seconds, unless terminated within that period by a decision response) and the length of the interval between stimulus termination and the next presentation (constant at 2 seconds). It also measured the interval between stimulus onset and the primary decision response, classified both decision and confidence responses, and stored a record of these statistics, summaries of which were printed out immediately after the training trials prior to the main session and again after the main session had elapsed.

Observers depressed either the left or right of two perspex buttons, lying on either side of a wedge-shaped resting position, to indicate whether the left or right line was judged longer. Immediately after this response one of a line of ten buttons, divided in the middle by a yellow line so that five were in front of each primary decision button, was depressed to indicate degree of confidence. For both judgements the confidence scale ran from

the button lying closest to the yellow line ("least confidence") to the outermost button ("most confidence").

(c) Design

An experimental session consisted of a block of 50 training trials followed by 1000 test trials. During the training trials the longer line appeared on the left and on the right in random order, but was equiprobable on either side. The test session contained 20 blocks of 50 trials each. The probability that the longer line would occur to the left or right was constant within each block, but varied between blocks according to the particular experimental condition. No gaps occurred between blocks so that observers would monitor an unbroken sequence of events.

Within each condition there were three main test phases: an initial period of constant signal probability, a period during which signal probability declined, and a final period at a constant lowered signal probability. (The term "signal" here refers to the stimulus in which the longer line appeared on the right. It was chosen arbitrarily for purposes of data classification and analysis. No such meaning was attached to either stimulus in the training of observers.) In condition A(i) the longer line occurred to the right and to the left on 25 trials during each of the first four blocks. The number of trials in which it appeared on the right was then progressively reduced to 20, 14 and 8 during the next three blocks. Over the final thirteen blocks it occurred on 3 trials during each block. Condition A(ii) was the same as A(i) except that the first phase was extended for an additional two blocks so that the probability of a longer line on the right declined after 300, rather than 200, trials. This second condition was added in order to determine whether additional experience at an "expected" signal probability would result in a greater or lesser rate of adaptation in response to change in that probability.

In condition B the longer line appeared on the right on 45 trials during each of the first four blocks (an increase in signal probability above its level in practice). This was progressively reduced to 37, 29, 21 and 13 appearances during the next four blocks. Over the final twelve blocks it occurred on 5 trials per block.

In all three conditions signal stimuli appeared randomly within each block so that no sudden change in relative probability would be perceived, with the likely exception of the initial block in condition B. Figure 20 shows the probability that the longer line would occur on the right during each of the 20 trial blocks.

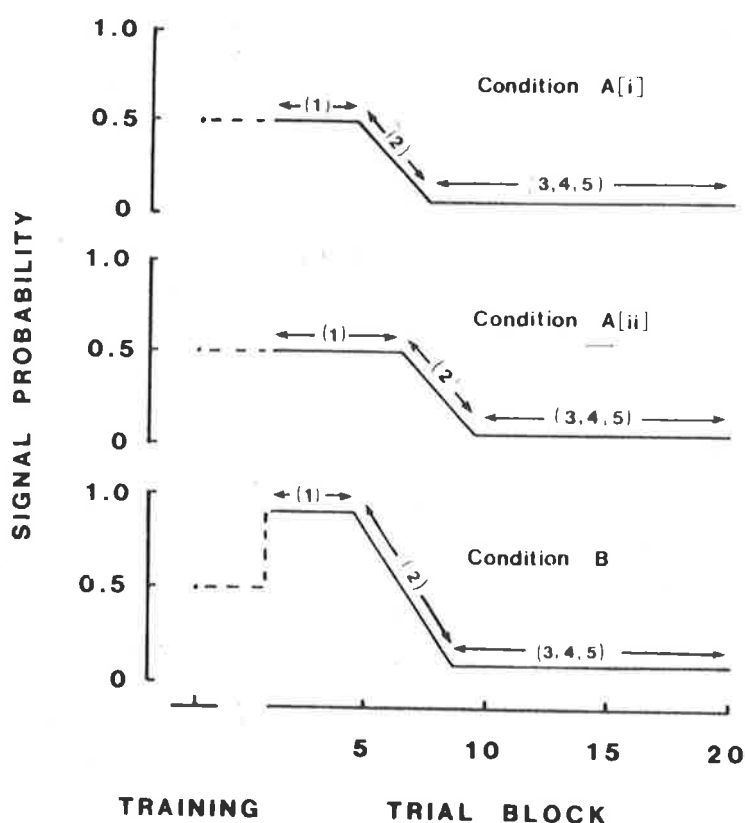


FIGURE 20 The probability that the longer line would occur to the right (signal probability) during training and each of the twenty successive trial blocks. In each condition five consecutive temporal periods were defined which are indicated by the bracketed numbers within each segment of the graphs.

(d) Procedure

Observers were tested individually and in complete isolation. Watches were removed before entering the experimental chamber. Prior to the training trials instructions were given which explained the nature of the discrimination task, the use of the response buttons, and the confidence scale. It was stressed that both the initial decision response as well as the subsequent confidence rating had to occur within 4 seconds of stimulus onset, and that the index finger of the dominant hand was to be used to make all responses. Observers were told that during the training trials the longer line would occur with the same frequency on either side of the figure so that approximately 25 could be expected on each side. Following training the print-out of performance data was checked to determine whether the observer was responding correctly. After this check observers were told that their performance was satisfactory and to continue to respond in the same manner during the main session to follow. No further reference was made of relative stimulus probabilities.

(c) Observers

Thirty-six females enrolled in first-year Psychology at the University of Adelaide volunteered to participate in the experiment as part of their course requirements. They were randomly divided into three groups of 12, each group to be tested under one experimental condition. Ages ranged from 17 to 20 years, with a median age of 18 years. All observers had normal or corrected vision.

2. Results

Prior to analysis, the data of the 1000-trial session were classified according to experimental condition, the signal probability phase within each

condition, and temporal period. There were three phases of signal probability corresponding to the initial phase of constant probability, the phase of intermediate probability during which signal probability declined, and the final phase of constant lowered probability. In addition, five temporal periods were defined (see Figure 20) where the first was identical to the initial phase of constant probability and the second corresponded to the second phase in which the probability of a signal declined. Within each condition the longer third phase was partitioned into three approximately equal divisions in order to obtain the final three periods. This allowed tests for change in performance at constant signal probability over the longer final phase of the session.

(a) Response probability

(i) Conditional signal response probabilities. The results of analyses of variance on response probability data are summarized in Table 5. The significant main effects of signal probability on the probability of correct and incorrect signal responses can be seen in Figure 21. There was an increase in both the probability of a correct signal report and the probability of a false signal report as signal probability decreased. In contrast, further analyses of the probability of a correct signal response over the final three time periods yielded no significant effect due to treatment condition, time period, or their interaction ($p > .3$ in all cases). However, the increase indicated in Figure 21 in the probability of an incorrect signal response, or false report, during this phase of the experiment was significant ($F = 17.727$, d.f. = 2/66, $p < .00001$), while there was no main effect due to condition ($p > .1$) or their interaction ($p > .8$).

TABLE 5. Results of analyses of variance on response probabilities.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
<u>(a) Correct signal response $P(S/s)$</u>					
	Condition (C)	2	.123	1.846	
	S within C	33	.066		
	Signal Probability (SP)	2	.219	14.660	.00001
	C x SP	4	.012	.832	
	SP x S within C	66	.014		
<u>(b) Incorrect signal response $P(S/n)$</u>					
	Condition (C)	2	.292	1.863	
	S within C	33	.156		
	Signal Probability (SP)	2	1.825	85.386	.00001
	C x SP	4	.038	1.807	
	SP x S within C	66	.021		
<u>(c) Signal response $P(S)$</u>					
	Condition (C)	2	.208	2.237	
	S within C	33	.093		
	Signal Probability (SP)	2	5.555	203.903	.00001
	C x SP	4	.711	26.114	.00001
	SP x S within C	66	.027		
<u>(d) Correct response $P(S/s)$ and $P(N/n)$</u>					
	Condition (C)	2	.122	.914	
	S within C	33	.133		
	Signal Probability (SP)	2	1.339	69.555	.00001
	C x SP	4	.112	5.857	.00005
	SP x S within C	66	.019		

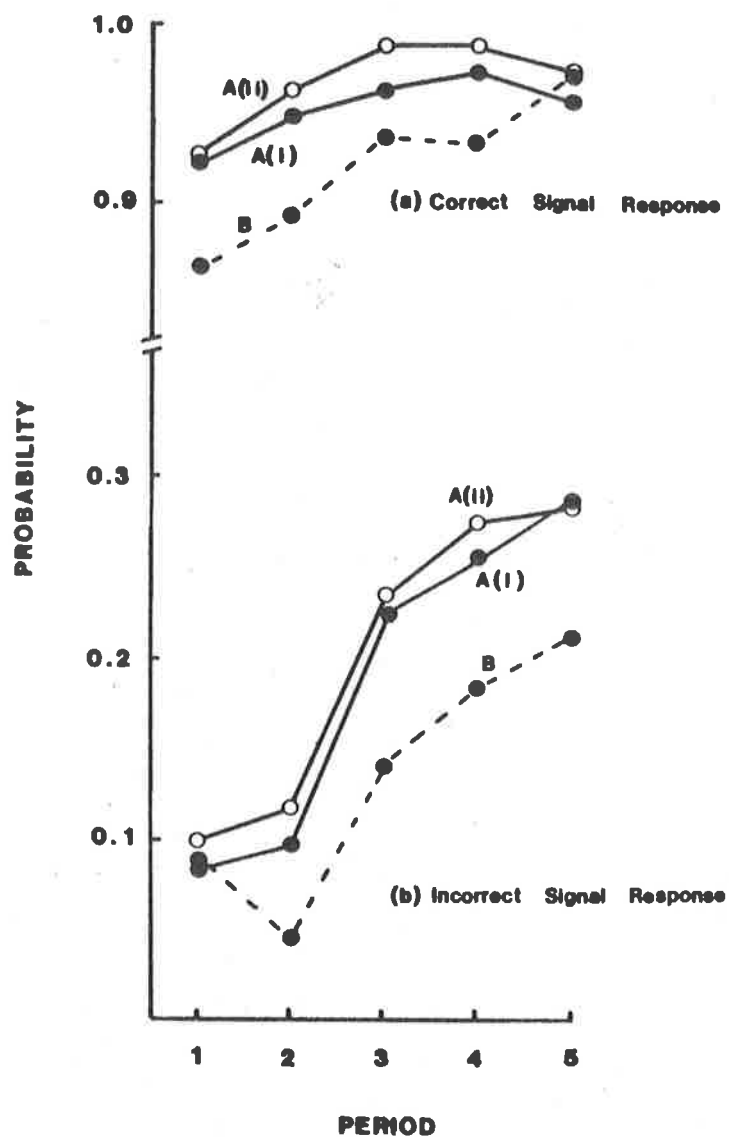


FIGURE 21. Changes in the mean conditional probability of (a) correct and (b) incorrect signal responses over the five temporal periods within each experimental condition.

(ii) Unconditional signal response probability. The significant main effect of signal probability and its interaction with treatment condition on the unconditional probability of making a signal response is apparent in Figure 22. As would be expected, a decline in the probability of making a signal response accompanied the drop in signal probability in all conditions. The interaction supports the proposal that adjustments made in response to

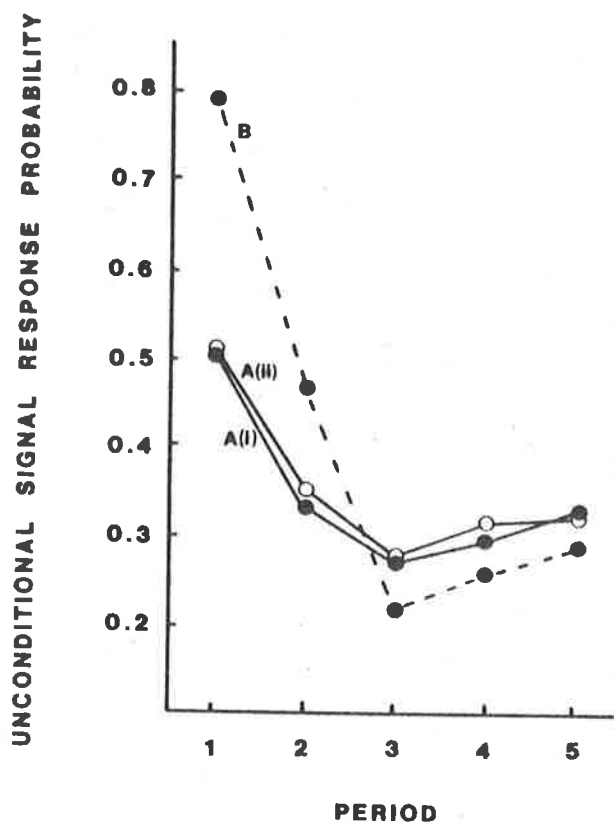


FIGURE 22. Changes in the mean unconditional probability of making a signal response over the five temporal periods within each experimental condition.

probability change are proportional to the initial discrepancy between L and C . At the beginning of the session observers in condition B made more signal responses than those observers in the two A conditions, as the actual signal probabilities would warrant. After signal probability declined these observers then made fewer signal responses than those in the A conditions, even though the signal probability under which they monitored was still higher (.10 versus .06). As the difference between the initial and final level of signal probability, and therefore the magnitude of $|L-C|$, was much greater in condition B, such larger adjustments would be expected by the adaptive model. The general increase in signal response probability over the final phase of the experiment, as indicated in Figure 22, was also

significant ($F = 17.048$, d.f. = $2/66$, $p < .0001$), although neither the effect of condition nor the condition $\times$ period interaction was significant ($p > .4$ in both cases).

(iii) Correct response probability. An interaction with experimental conditions also accompanied the significant main effect of signal probability on the unconditional probability of making a correct response, signal and nonsignal responses combined. Figure 23 shows the effect to be a decline in correct response probability as signal probability declined, the degree of decline being the least in condition B. The lower error rate of condition B during the final phase of the experiment occurred in conjunction with a lower probability of making a signal response. Thus the increase in error

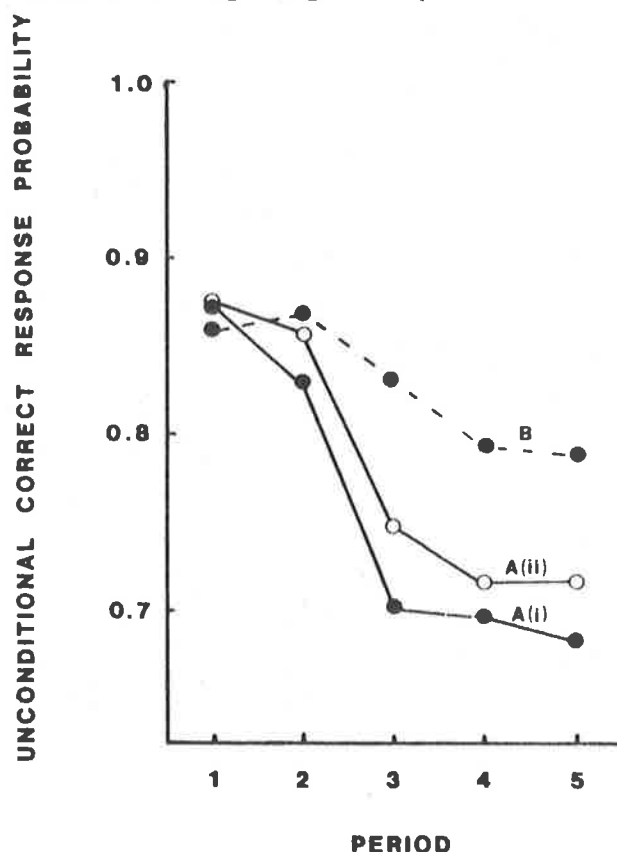


FIGURE 23. Changes in the mean unconditional probability of making a correct response (signal or nonsignal) over the five temporal periods within each experimental condition.

rate, produced by disproportionate increases in false report rates relative to correct detection rates (see Figure 21), was greatest in the A conditions where the total proportion of signal responses was higher. The tendency for the proportion of correct responses to decline over the final phase of the experiment was significant ($F = 3.549$, d.f. = 2/66, $p < .05$). The differences shown between conditions in the proportion of correct responses during this phase of the experiment was of borderline significance ($F = 2.894$, d.f. = 2/33, $p = .07$), while the interaction between condition and time period was nonsignificant ($F = .363$, d.f. = 2/66, $p > .8$).

(b) Response latencies

(i) General trends. Owing to the small number of false reports during the initial phase of the session, and missed signals during the final phase, in all three conditions, no reliable analyses could be carried out on the reaction times of the two types of incorrect response. The available data were generally quite sporadic and no consistent trends were evident. Analyses of variance on the latencies of the two types of correct response are summarized in Table 6. There was a significant main effect of signal probability on both correct signal and correct nonsignal response latencies. Correct signal responses tended to be made more quickly as signal probability decreased and Figure 24 indicates that this trend was more pronounced in the A(ii) and B conditions, although there was no significant interaction. Corresponding to the decrease in correct signal response latency there was a tendency for correct nonsignal response latency to increase as signal probability declined. There was no significant effect on the latencies of either response over the final three watch periods ($p > .3$ in all cases).

TABLE 6. Results of analyses of variance on response latencies.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
<u>(a) Correct signal response</u>					
	Condition (C)	2	.197	.156	
	S within C	33	1.266		
	Signal Probability (SP)	2	.443	4.511	.05
	C x SP	4	.126	1.285	
	SP x S within C	66	.098		
<u>(b) Correct nonsignal response</u>					
	Condition (C)	2	2.825	1.469	
	S within C	33	1.922		
	Signal Probability (SP)	2	1.764	13.462	.00001
	C x SP	4	.148	1.136	
	SP x S within C	66	.131		

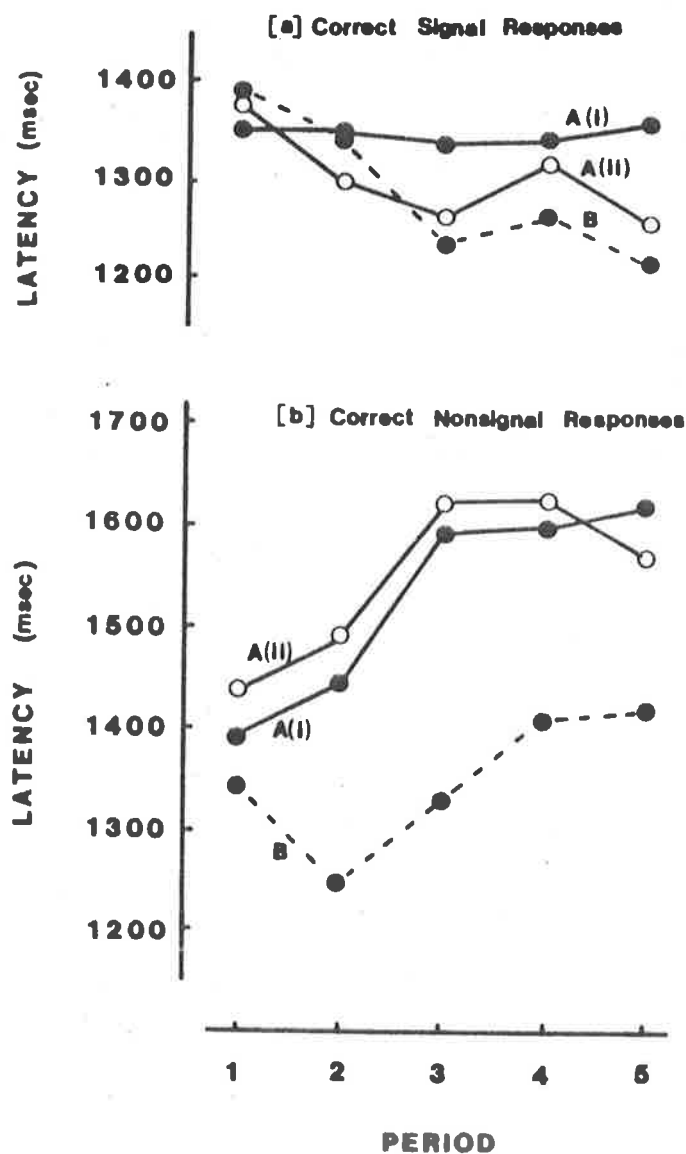


FIGURE 24. Changes in the mean latencies of (a) correct signal and (b) correct nonsignal responses over the five temporal periods within each experimental condition.

(ii) Differences among the four response categories. Table 7 contains the mean latency for each response category within each condition. Differences between each pair were subjected to paired-sample t-tests. In every condition correct signal reports were made more quickly than incorrect reports of signals (A(i) : $t = -14.97$, d.f. = 216, $p < .001$; A(ii) : $t = -14.56$, d.f. = 208, $p < .001$; B : $t = -10.14$, d.f. = 176, $p < .001$), correct nonsignal reports were made more quickly than incorrect nonsignal reports (A(i) : $t = -5.36$, d.f. = 57, $p < .001$; A(ii) : $t = -2.85$, d.f. = 57, $p = .003$; B : $t = -6.01$, d.f. = 103, $p < .001$), and while differences between the two types of incorrect response were not significant (A(i) : $t = .60$, d.f. = 46, $p = .55$; A(ii) : $t = -1.06$, d.f. = 48, $p = .29$; B : $t = .98$, d.f. = 53, $p = .33$) correct signal responses were made more quickly than correct nonsignal responses (A(i) : $t = -9.78$, d.f. = 238, $p < .001$; A(ii) : $t = -12.19$, d.f. = 239, $p < .001$; B : $t = -3.60$, d.f. = 239, $p < .001$). The first two findings agree with the distance-from-criterion interpretation of response latencies, as discussed in Chapter 1, in which it is argued that observations lying

TABLE 7. Mean latencies and confidence for the four categories of response.

<u>Condition</u>	<u>(a) Response latency (msec)</u>			
	<u>Correct Signal</u>	<u>Incorrect Signal</u>	<u>Correct Nonsignal</u>	<u>Incorrect Nonsignal</u>
	(S/s)	(S/n)	(N/n)	(N/s)
A(i)	1347	1707	1537	1784
A(ii)	1312	1696	1539	1544
B	1290	1522	1350	1556
	<u>(b) Response confidence (5-category scale)</u>			
A(i)	3.15	2.03	2.45	1.81
A(ii)	3.18	2.04	2.58	1.92
B	3.28	2.43	3.04	2.03

close to the decision criterion give rise to more difficult discriminations and so take longer to result in a decision than observations lying further from the criterion. Hence for both decision alternatives correct reports should be made faster than their incorrect counterparts. The latter result, shorter correct signal than correct nonsignal response times, then implies that the decision criterion has moved leftward along the nonsignal, or noise, distribution. The increases in both conditional signal response probabilities, and accompanying decreases in correct signal response latencies and increases in correct nonsignal response latencies, also suggest a decline in the criterion used in making a signal response. This opposes those trends in response times found in conventional detection studies where a decline in signal probability led to longer signal response latencies and shorter nonsignal response latencies (e.g. Carterette, Friedman & Cosmides, 1965; Gescheider, Wright & Evans, 1968; Gescheider, Wright, Weber, Kirchner & Milligan, 1969), presumably due to an increase in the critical value of the criterion for a signal response.

(c) Response confidence

(i) General trends. For the reasons mentioned above in section (b), no analyses could be performed on the confidence data for incorrect responses. As with response latencies no consistent trends were observable in the data. Table 8 contains the results of analyses of variance on confidence in the two types of correct decision. The significant effect of signal probability and its interaction with condition on correct signal response confidence is evident in Figure 25. Confidence increased as signal probability declined in conditions A(i) and B but showed no change in condition A(ii). This is the only difference in performance trends found between the two A conditions. Thus the only effect of extending the number of trials at the initial "expected" level of signal probability was to suppress the tendency for confidence in

TABLE 8. Results of analyses of variance on response confidence.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a) Correct signal response					
	Condition (C)	2	1.166	.074	
	S within C	33	15.748		
	Signal Probability (SP)	2	10.756	12.729	.00005
	C x SP	4	4.534	5.366	.001
	SP x S within C	66	.845		
(b) Correct nonsignal response					
	Condition (C)	2	23.038	1.692	
	S within C	33	13.610		
	Signal Probability (SP)	2	14.304	14.289	.00001
	C x SP	4	1.684	1.682	
	SP x S within C	66	1.001		

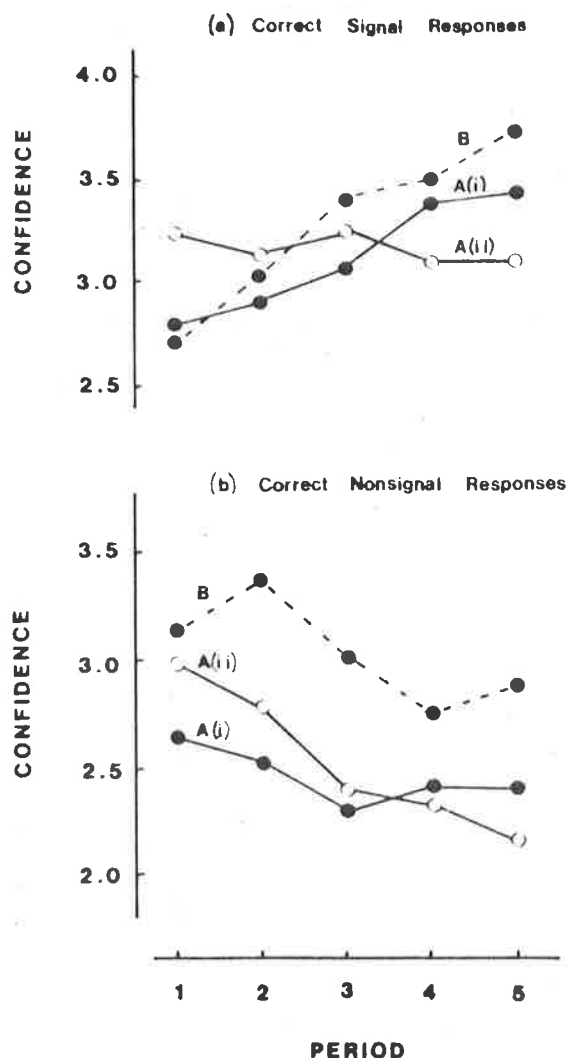


FIGURE 25. Changes in the mean confidence expressed in (a) correct signal and (b) correct nonsignal responses over the five temporal periods within each experimental condition.

correct signal reports to increase as signal probability declined. Also evident in Figure 25 is the significant main effect of signal probability on confidence in correct nonsignal responses. There was a decrease in confidence to accompany the decrease in signal probability in all conditions. The sole effect to emerge from analyses of change over the final three watch periods was an interaction between treatment condition and time period on confidence in correct signal responses, which reached marginal significance ($F = 2.422$, d.f. = 4/66, $p = .057$). While confidence appeared to increase in the A(i) and B conditions (Figure 25), a slight downward trend occurred in condition A(ii).

(ii) Correspondence to results obtained with response latencies. The above results are consistent with the inverse relationship shown to exist between response latency and level of confidence (e.g. Katz, 1970; Colquhoun & Goldman, 1972; Emmerich, Gray, Watson & Tanis, 1972) which was interpreted in Chapter 1 in terms of the distance-from-criterion hypothesis (see Figure 2). As correct signal report reaction time decreased with the drop in signal probability, the confidence in that report increased, and vice versa for correct nonsignal reports. Further, differences between confidence means within conditions were all opposite in sign to those differences found between latency means, as can be seen in Table 7. As with response latencies, paired-sample t-tests were carried out on these differences. In all conditions the same response pairs showed significant differences in mean confidence as had exhibited latency differences, the differences now being of the opposite sign: more confidence was expressed in correct signal reports than in incorrect signal reports (A(i) : $t = 18.95$, d.f. = 216, $p < .001$; A(ii) : $t = 17.59$, d.f. = 208, $p < .001$; B : $t = 16.74$, d.f. = 176, $p < .001$), correct nonsignal reports were made more confidently than incorrect nonsignal reports (A(i) : $t = 8.75$, d.f. = 54,

$p < .001$; A(ii) : $t = 9.09$, d.f. = 56, $p < .001$; B : $t = 10.15$, d.f. = 102, $p < .001$), and while the difference in the confidence expressed in the two types of incorrect response was not significant (A(i) : $t = -.76$, d.f. = 43, $p = .45$; A(ii) : $t = -.18$, d.f. = 48, $p = .86$; B : $t = .07$, d.f. = 52, $p = .94$) correct signal reports were made more confidently than correct nonsignal reports (A(i) : $t = 12.63$, d.f. = 238, $p < .001$; A(ii) : $t = 14.07$, d.f. = 239, $p < .001$; B : $t = 4.36$, d.f. = 239, $p < .001$).

(d) Signal detection measures and theoretical decision parameters

(i) Effects of signal probability. The changes in response probability, latency and confidence all serve to suggest that the main effect of the decrease in signal probability was a shift in bias toward the signal response alternative. Freeman's (1964) tables of d' and β were used to convert conditional signal response probabilities into signal detection measures and so test this assertion. The rarity of missed signals after the decline in signal probability made a thorough analysis of change in these measures impossible. It was decided not to use conventions such as assigning a missed signal, or the probability of a missed signal had the entire group of observers missed only one signal, to the data of an individual. This would have been particularly undesirable in an analysis of change over the last three watch periods since most observers would have had 'corrected' data for at least two of the three periods. However, values of d' and β were calculable at the initial and final levels of signal probability for eight observers in condition A(i) and six observers in condition B. In condition A(ii) measures could be derived for only two observers. Therefore, analyses of variance were performed on the data of the observers from the A(i) and B conditions only, and the results are given in Table 9. It should be noted that there were no differences in the trends of the other performance measures between those

TABLE 9. Results of analyses of variance on signal detection measures.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	Sensitivity (d')				
	Condition (C)	1	1.268	2.465	
	S within C	12	.514		
	Signal Probability (SP)	1	.363	3.845	
	C x SP	1	.363	3.845	
	SP x S within C	12	.094		
(b)	Bias (β)				
	Condition (C)	1	.541	.917	
	S within C	12	.590		
	Signal Probability (SP)	1	4.546	8.344	.05
	C x SP	1	1.522	2.794	
	SP x S within C	12	.544		

observers whose conditional response probabilities could be converted into signal detection measures for the analyses, and those who did not produce enough misses or false alarms to permit the measures to be calculated at the two levels of signal probability. The subgroups whose data entered into the analyses can therefore be considered as representative of the groups from which they were drawn, except for the fact that they were slightly less accurate in their responding.

The fact that the effect of signal probability was upon β rather than d' (Table 9) is in accord with the argument put forth in the previous paragraph. Moreover, Figure 26 shows that the main effect of decreasing signal probability was a decrease in β , rather than the increase predicted by the response optimization model. On the other hand, the response stabilization model would predict a fall in response to the negative value of ($L-C$) produced by the changes in local probability, and a subsequent tendency for β to increase when the value of ($L-C$) tended toward zero. Observers in the present study would have based any decisions concerning the probabilistic

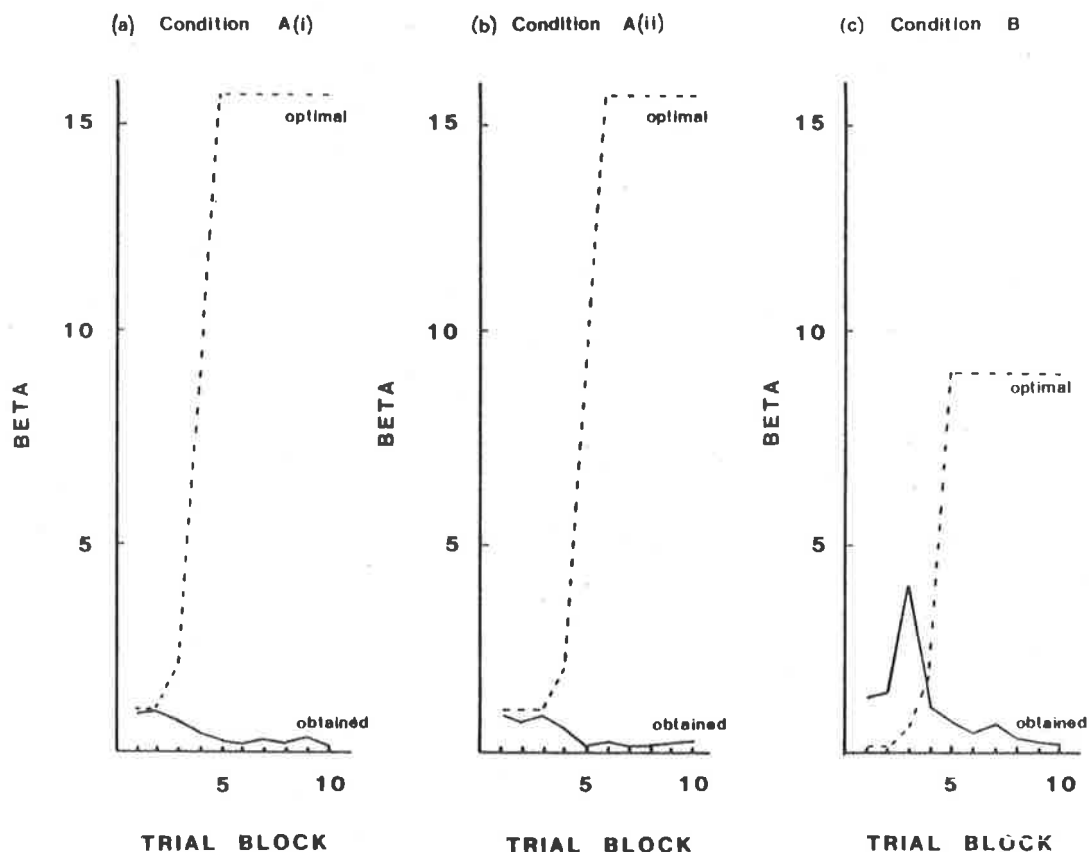


FIGURE 26. Changes in group values of β over successive blocks of 100 trials during the experimental session. In the figure for each condition the values of β that were observed are compared with the optimal values specified by signal detection theory.

nature of the stimuli on response, rather than the actual stimulus, probabilities. Changes in β were therefore compared with changes in the statistics, l and c . These are the respective local and cumulatively-based probability of making a signal report. Changes in $(l-c)$ should have the same effect on β as corresponding changes in $(L-C)$, the difference between the objective probabilities.

(ii) Relationship between β and $(1-c)$. Figure 27 shows change in the mean value of $(1-c)$ calculated for each of the twenty trial blocks in each condition. Practice trials were used to calculate the initial, as well as subsequent, values of c . The finest-grain analysis that could be made of

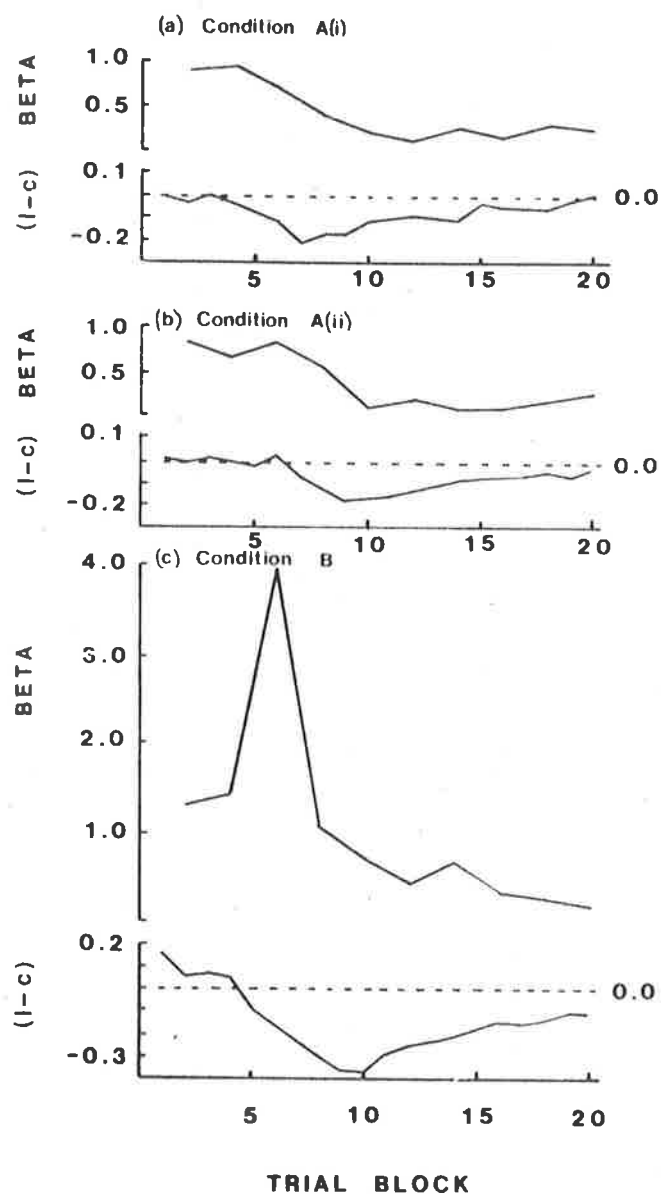


FIGURE 27 A comparison of changes in group values of β and mean values of $(1-c)$ during the experimental session in each condition. Values of β were computed over ten successive blocks of 100 trials each, and $(1-c)$ over twenty successive blocks of 50 trials each.

β , without the conventions for increasing missed signal probability, was a plot of values calculated from group means of conditional signal response probabilities for successive blocks of 100 trials. These values, which are also those shown in Figure 26, were then used for comparison with $(1-c)$ in Figure 27. Conditions A(i) and A(ii) conform well with the predictions of the response stabilization model. At constant signal probability in the beginning of the session $(1-c)$ modulated around the zero point and there was no change in β . Then as $(1-c)$ became negative when signal response probability declined β showed a decline which, at least for condition A(i), appeared to continue past the maximum negative value of $(1-c)$. Finally, as $(1-c)$ tended towards, and in Condition A(i), reached, zero, β showed an upwards trend. The relationship between the measures in condition B is less straightforward. The model predicted that β would have an initially high value which would decline and then, toward the end of the session, increase. Although β was initially higher than in the A condition, its maximum was reached during the fifth and sixth blocks of trials when $(1-c)$ had already declined past the zero point. It then showed a large decline during the remainder of the decline in $(1-c)$, but continued to decline beyond the point where $(1-c)$ had begun to return toward zero, and was still declining at the end of the session.

The response stabilization model demands a strong positive relationship between β and $(1-c)$. Therefore, in order to statistically test how well the above changes in β within each condition could be considered to correspond to changes in $(1-c)$, Pearson product-moment correlation coefficients were calculated between the two measures. As the finest computations that could be made of β were of group values calculated from mean conditional response probabilities over blocks of 100 trials, these values

were compared with mean values of $(1-c)$ also calculated over successive blocks of 100 trials. Two separate correlations were tested in each condition: the first ($n = 10$) between $(1-c)_i$ for any trial block, i , and the value of β in that trial block, β_i , and the second ($n = 9$) between $(1-c)_i$ and the value of β in the following trial block, β_{i+1} . The second analysis was performed since, although any appreciable change in $(1-c)$ will demand a compensatory change in β in the same direction according to the stabilization model, the change in $(1-c)$ must first be detected and then acted upon by the stabilization mechanism, so that change in β must necessarily follow that in $(1-c)$ by some finite period of time.

As can be seen in Table 10, when $(1-c)$ was compared with β from the same trial block only the positive correlation found in condition A(ii) reached significance, although the correlations in the other two conditions were both positive. However, when $(1-c)$ was compared with the value of β which occurred in the next trial block there were significant positive correlations obtained in all three conditions. Hence, despite the lack of

TABLE 10. Pearson product-moment correlations between the mean value of $(1-c)_i$ in each block of 100 test trials and the group value of β , either in that trial block (β_i) or the following trial blocks (β_{i+1}).

<u>Condition</u>	<u>Correlation Coefficient</u>	
	$(1-c)_i$ with β_i	$(1-c)_i$ with β_{i+1}
A(i)	.329	.680 *
A(ii)	.762 *	.833 *
B	.245	.677 *

* $p < .05$

perfect coincidence of changes in the two measures, they are highly related, particularly when the time taken for the stabilization mechanism to intervene and act upon any alteration in the response ratio is taken into account. This "lag" might also explain the seemingly weaker relationship mentioned above to be shown between β and $(1-c)$ in condition B. The correlation between the two measures increased the most in this condition when $(1-c)$ was compared with β from the following trial block rather than the same trial block (Table 10). Hence, the effect of any processing interval may have been greatest in this condition due to the larger changes in objective probabilities and the more pronounced criterion adjustments which they specified. As shown in Figure 19(b), the response stabilization model demanded an initial increase in the decision criterion followed by two periods during which the criterion decreased and a further, and final, increase. Figure 27(c) indicates the first three of these trends took place, but later in the event sequence than was predicted by the model based on objective probabilities. It is then possible that if the length of the session had been extended the final predicted increase in criterion may have been displayed. Alternatively, there may have been no need for the final adjustment. That is, the somewhat dramatic alterations in the relationship between the values of l and c which occurred in the earlier stages of the task may have led to a larger discrepancy between l and c being tolerated later in the task.

3. Discussion

(a) Relative support for the three models

(i) Relationships among performance measures. In each condition as the conditional probabilities of making a signal response increased, there was a decrease in the latency of, and increase in the confidence expressed in, correct signal reports. At the same time correct nonsignal reports were made more slowly and with decreased confidence. These trends are in agreement with those that would be predicted on the basis of the distance-from-criterion hypothesis of signal detection (Chapter 1, section B.4.(d)) to follow upon a decline in the critical value of the criterion for a signal response. Further support stems from the finding that, for both signal and nonsignal reports, incorrect responses took more time and resulted in less confidence than correct responses. Finally, consistent with a critical value of the decision criterion of less than 1.0, as was observed during the greater part of the task, correct signal reports were made more quickly and confidently than correct nonsignal reports.

(ii) Main effect of level of signal probability. Despite the support gained for the distance-from-criterion hypothesis, it is evident that when observers encounter a change in signal probability they do not behave in the fashion specified by the response optimization model of signal detection. This model predicted an increase in β to follow upon the decrease in signal probability. Instead, the changes in response probability, latency, and confidence measures all coincided with a decrease in the observed value of β . Therefore observers did not alter their decision criterion in the optimal fashion specified by the change in the value of $p(n)/p(s)$, the ratio of the probabilities of nonsignal and signal stimuli.

The decrease in β is consistent with the predictions of both the response maintenance and the response stabilization models. However, the trends observed in the other response measures give clear evidence that response maintenance was not taking place. In the model outlined in Chapter 2 (section B.3.(c).(ii)) it was shown that following a decline in signal probability response maintenance could be achieved only through a disproportionately large increase in false alarm rate relative to detection rate and would be reflected by signal detection analysis as decreases in both d' and β . These trends were not observed in the present study. For instance, when the data of the two A conditions are combined, as the probability of a correct detection increased from .925, over the initial phase where signal probability was set at .5, to .972, over the final phase where signal probability was .06, the probability of a false report increased from .091 to .260. However, using the strategy of response maintenance given in Chapter 2, a considerably larger increase in false report rate, to approximately .48, would have been expected. Further, while d' showed only a slight decline, from 2.775 to 2.559, β declined from .864 to .198, a rather large change for a likelihood ratio measure.

(iii) Performance change over the final phase of constant signal probability. Only the response stabilization model predicted any change in performance over the final phase of the task where signal probability remained constant at that value to which it had been lowered. This prediction was based on the fact that the absolute difference between local and cumulatively-based signal probability was here decreasing. Consistent with the model, after $(1-c)$ in the two A conditions had reached its maximum negative value and then tended to increase toward zero, there appeared to be an increase in the decision criterion for a signal response. Although the increases in

the value of β were small, and could not be tested statistically, a more striking picture emerges if changes are compared in values of $(1-c)$ and β based upon the conditional and unconditional probabilities of the "nonsignal" response. These values can be calculated since the design of the present task allowed either of the two stimuli to be classified as the signal stimulus for the purpose of analysis. When the stimulus with the longer line to the left is treated as the signal in this manner, the changes in the likelihood measure, β , become more pronounced since β is now the inverse of values between 0.0 and 1.0. Figure 28 shows the result of this presentation of the data. After $(1-c)$ declines toward zero in condition A(i) and A(ii) there are quite evident declines in the value of β , as would be expected by the response stabilization model.

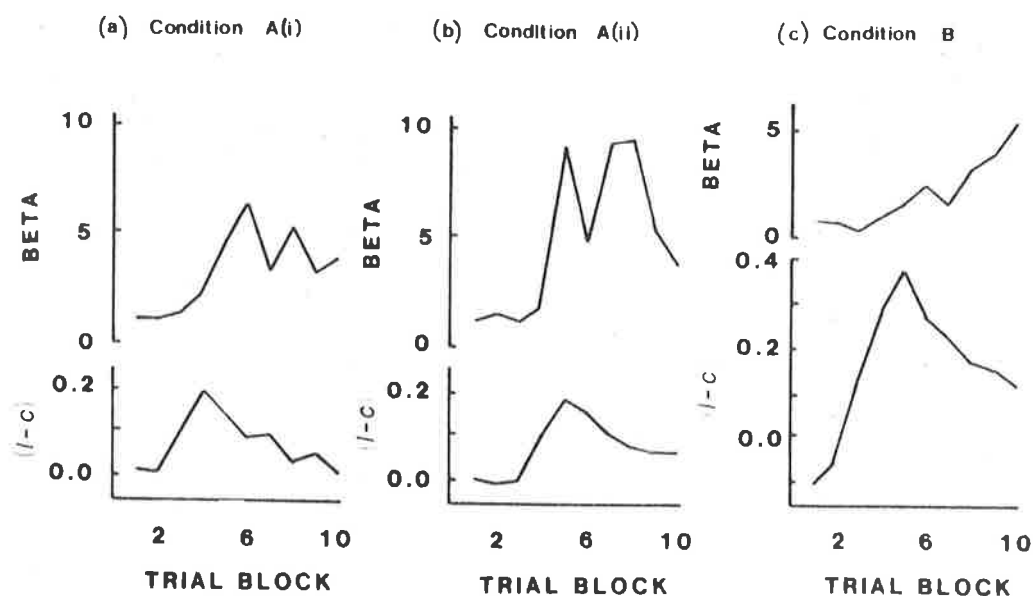


FIGURE 28 A comparison of changes during the session in group values β and mean values of $(1-c)$. In each condition the values of β and $(1-c)$ are here based upon the mean probabilities of reporting the stimulus with the longer line to the left in each of ten successive blocks of 100 trials each.

Figure 28 also heightens the discrepancy in condition B between the change observed in the value of β over the final phase of the task, and that predicted on the basis of change in $(1-c)$. The continued upward trend in β beyond the point where $(1-c)$ started to decline now appears too pronounced to be considered the result of a larger discrepancy being tolerated between l and c . Also, since it persisted for 500 trials beyond the point at which $(1-c)$ reached its maximum, it is unlikely that it is the result of the decrease in the value of $(1-c)$ not being noticed by the stabilization mechanism due to the number of trials which must elapse before estimates of L and C are revised and intervention in the decision process occurs. It is possible that, after making the initial adjustments specified by the mechanism, observers at this later stage of the task were adopting a strategy of altering their decision criterion in such a way that would simply increase (or decrease) the local probability of a signal response toward the cumulatively-based probability. This would still satisfy the goal of the stabilization model, the minimization of any discrepancy between local and cumulatively-based signal probability, and it can be seen from Figures 27 and 28 that $(1-c)$ did tend toward zero in condition B in the same manner as in the two A conditions.

The argument that observers in condition B may have adopted such a strategy later in the task, however, does not mean to imply that all observers were behaving in the fashion specified by the hypothesis put forward at the end of Chapter 2 (section B.3.(b)). That is, they were not always adopting a value of the decision criterion which would result in a matching of local and cumulatively-based signal probability. In the two A conditions local signal probability was always less than cumulatively-based signal probability after the initial decrease in signal probability. A constant decline in the value of β would have been required in order to

increase the local probability of a signal response toward cumulatively-based signal probability. This was not found in either condition. Also, the results cannot be accounted for by assuming that they reflected a tendency for response ratios to successively approximate stimulus ratios. Whereas the local probability of a signal stimulus was always less than that of a signal response over the final phase of the experiment, the unconditional probability of a signal response increased during this portion of the task in all conditions. Hence, the alterations in the decision criterion demanded by the stabilization model provides the most adequate explanation for the majority of findings in the present tasks, but there is some evidence that the observer may switch to different response strategies depending upon the nature of the task.

(b) Relevance of present findings to vigilance

The response stabilization model was advanced originally in an attempt to explain why vastly different trends may occur in response bias during the course of vigilance tasks. However, despite the degree of support gained for the model, the generalizability to the area of vigilance of any conclusions based on the results of the present experiments can be questioned. This is due to the nature of the discrimination task which has been employed, as compared to the detection task of the traditional vigilance study where the signal typically represents an increment in magnitude above the level of nonsignal stimuli along some common dimension such as length, brightness, or loudness. There the task of the observer is to make a judgement on each trial as to whether the observed magnitude of sensory effect, x , was greater than that critical amount, x_c , necessary to make a signal response. The distributions of the sensory effect produced by signal, s , and nonsignal, n , stimuli are assumed by signal detection theory to lie along an observation axis as shown in Figure 29(a). (Section B.4.(a) of Chapter 1 outlines the

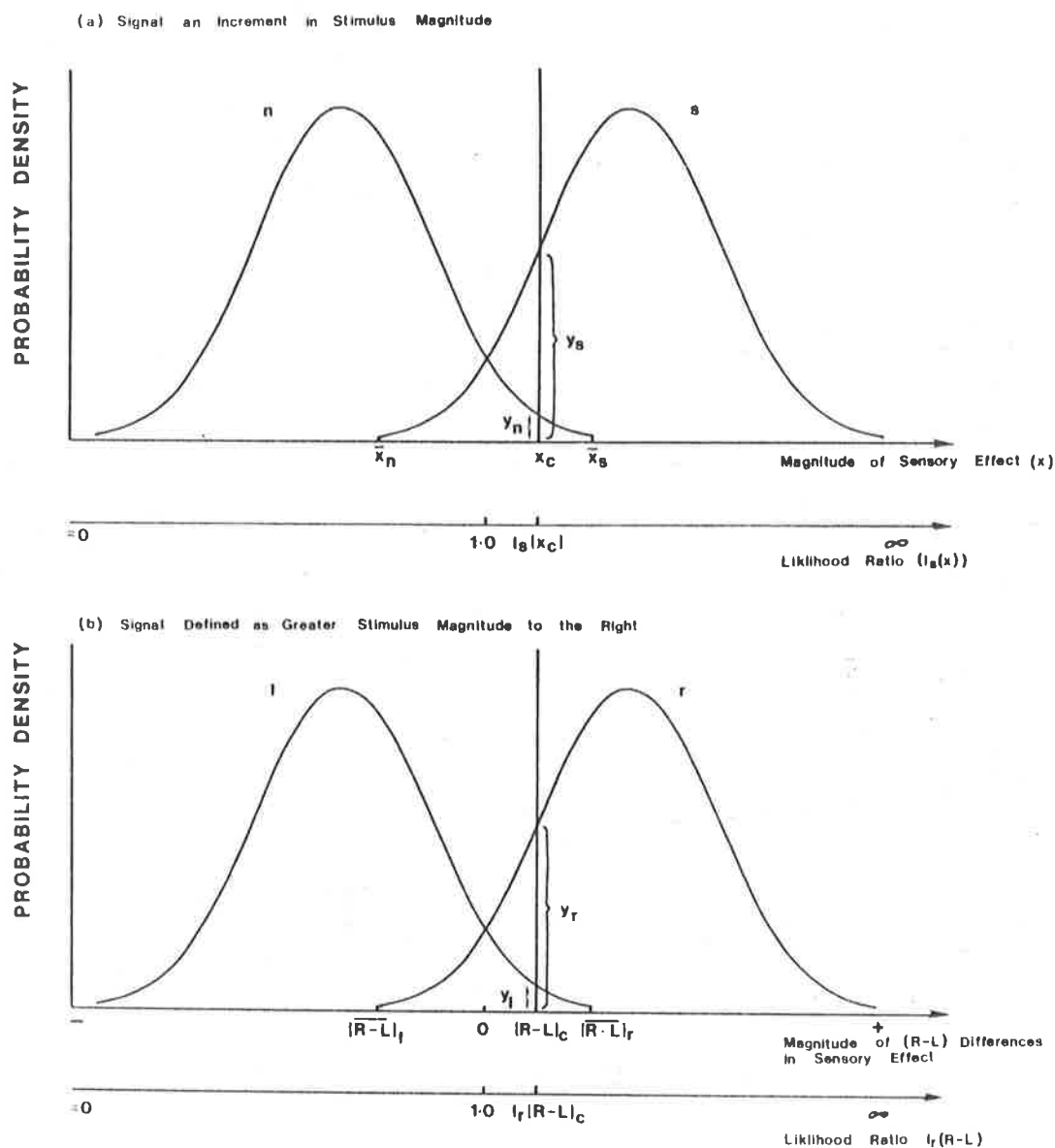


FIGURE 29. Hypothetical distributions of the sensory effect produced by signal and nonsignal stimuli when (a) the signal, s , is an increase in magnitude above the level of nonsignal, n , stimuli and (b) when the signal, r , is defined as the case when the stimulus to the right is of greater magnitude and the nonsignal, l , as the case when the stimulus to the left is of greater magnitude. In both (a) and (b) each value of the sensory variable (upper abscissa) is associated with a value of the likelihood that it arose from a signal presentation (lower abscissa), given by the ratio of the ordinate value, y , to the signal and nonsignal distribution, respectively. This ratio at the cutoff position for a signal response, β , is indicated for both types of signal.

major assumptions of the theory.) This axis is thus a continuum of the absolute magnitude of the sensory effect received during each observation interval.

In the present studies, however, observers had to make judgements about the relative lengths of two simultaneously presented lines, either of which would exceed the other by a constant amount on each trial. This situation can be treated in signal detection terms in the manner suggested by Treisman and Watts (1966), and further elaborated by Vickers, Nettelbeck and Willson (1972), by supposing that the observer makes a judgment based on the difference in sensory effect produced by the two lines. Specifically, the observer can be regarded as making an observation on each trial of the $(R - L)$ difference in the sensory effect produced by the line on the right, R , and that on the left, L . If on any trial the $(R - L)$ difference exceeds some critical value, $(R - L)_c$, then the observer will report the right to be greater, if it fails to do so the left will be reported as greater. Along the observation axis are now two distributions of sensory differences: that which results over time when the right stimulus is objectively greater, r , with the positive mean $(\overline{R - L})_r$, and that which results when the left is the greater, l , with the negative mean, $(\overline{R - L})_l$, as shown in Figure 29(b). These distributions can be assumed to be normal and of equal variances in the same manner as those envisaged in the traditional detection task.

There should be no difference in decision behaviour in the two types of task according to the classical theory of signal detection since the observer is assumed to transform the observation axis into a likelihood ratio decision axis. In Figure 29(a) each value of x is thus converted into the likelihood that it arose from the s distribution, $l_s(x)$, the ratio of the respective ordinate values to the s and n distributions at x . The value of the

likelihood ratio at the critical observation magnitude for a signal response, x_c , is then $l_s(x_c)$ or, as it is usually termed, β . On each trial that $l_s(x)$ exceeds β a signal response is made. Similarly, in Figure 29(b) "right greater" is taken to be the signal response and each $(R-L)$ difference in sensory effect then corresponds to the likelihood that it arose from the r distribution, $l_r(R-L)$, the ratio of the ordinate values to the r and l distributions, respectively, at the observation point, $(R-L)$ (Conversely, the ratio of the ordinates to the l and r distributions, respectively, would give the likelihood that the difference arose from l). As in the previous case, the value of the likelihood ratio at the critical magnitude of observed difference for a right response, $(R-L)_c$, is then $l_r(R-L)_c$, or β . If $l_r(R-L)$ exceeds the value of β on any trial the observer will report that the right is greater. The same sort of decision that underlies the report of an increase in sensory magnitude then applies to the judgment as to which of two sources is of the greatest sensory magnitude.

Alternative conceptions of the signal detection process may not predict such a correspondence in performance between the two types of task, particularly if the observer is not assumed to transform sensory activity into a likelihood ratio. For example, in the "neural timing" theory of Luce and Green (1972) the observer decides whether the length of the interval between successive neural impulses is less than some critical period, β , below which a signal will be reported. The hypothetical decision axis is then a continuum based directly on the rate of sensory activity, that is, the observation axis itself. A decision as to whether the present rate of neural activity represents an increase over some randomly varying background level, as would occur in a typical detection task, is essentially quite different from judging which of two separate sources is presently producing the greatest amount of activity. The latter decision would appear to be based

on more complex information, and could involve a completely different sort of decision rule. For instance, a response might be initiated as soon as the amount of activity from either source has reached some critical value, irrespective of the difference between the two, so that the decisions in the two situations are fairly similar. Alternatively, a response might take place only after the difference between the levels of activity of the two sources has reached some critical value, irrespective of the absolute levels of each, so that decision behaviour in the two situations is quite dissimilar. Moreover, the choice of decision rule would be based on the knowledge that there will be a difference in sensory input during each observation interval, whereas in the traditional detection task used in vigilance the observer must decide whether each observation interval in fact contained a difference in sensation from that associated with sensory noise alone. For these reasons, it was considered necessary to see whether the results obtained with the discrimination tasks could be replicated in a task which more closely resembled that typically employed in vigilance studies.

C. EXPERIMENT FOUR : EVALUATION OF THE RESPONSE

STABILIZATION MODEL IN A REPRESENTATIVE VIGILANCE TASK

In addition to the difference mentioned above in the characteristics of the traditional detection task of vigilance and the present discrimination tasks, a further discrepancy might arise from the manner in which observers are trained in the two situations. In the previous three experiments observers were given instructions in such a way that equal values were attached to the reports of both stimulus alternatives. This is in contrast to the procedure employed in vigilance tasks, where a common practice is to make a verbal distinction between signal and nonsignal stimuli which implies that more importance should be attached to one response than to the other. The initial degree of bias toward a signal response is undoubtedly affected by the use of terms such as "critical event" or "signal", and in drawing analogies between the task of the observer in the laboratory and in operational settings such as air traffic control or the spotting of defects on an assembly line. Hence it was the purpose of the present study to see whether those trends in performance that were observed in the two-category discrimination tasks would also be obtained in a signal detection situation where an initial asymmetry of the payoff matrix might exist. As in the previous experiment signal probability was declined in a ramp fashion and subsequently held constant. The response stabilization model again predicted a decrease in the criterion for a signal report during the phase of the task in which the local probability of a signal declined, followed by an increase during the final phase in which signal probability was held constant.

1. Method

(a) Task and stimuli

On each trial a 106 x 177 mm display consisting of horizontal lines which varied in length was presented on the screen of a television monitor and viewed by the observer at a distance of approximately 4.4 metres. The observer had to decide whether the display was a signal or nonsignal display. The nonsignal, or standard, display was one in which the horizontal lines varied in length but had a constant mean length throughout the display of 10.0 mm and range from 7.1 to 12.9 mm. In contrast, the lines in the central 84 by 106 mm matrix of the signal display had a mean length of 12.9 mm and a range of 10.0 to 15.8 mm. An example of each type of display is shown in Figure 30. Thus the task of the observer was to detect the presence of longer lines in the centre of the display.

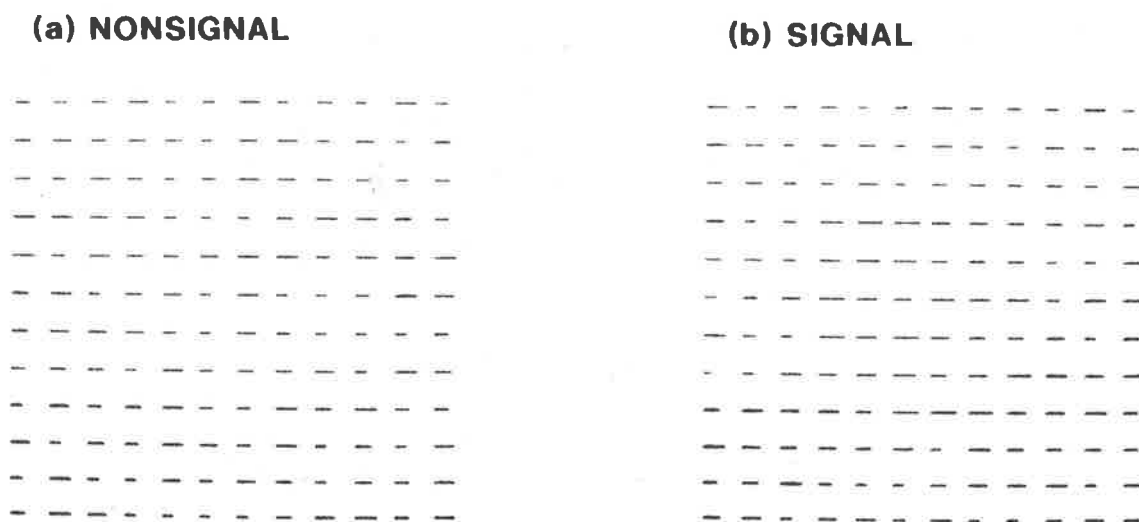


FIGURE 30. Examples of (a) a nonsignal and (b) a signal stimulus as used in Experiments 4 and 5. The average line length in the central matrix of the signal display is greater than the average line length in the perimeter. In the nonsignal display average line length is constant throughout.

Observers were required to respond when a nonsignal was judged to be present, as well as to make signal reports. Each type of response was followed by a confidence rating along a 5-category scale to indicate subjective certainty in the accuracy of the primary decision response.

(b) Apparatus

The graphics program contained in a D.E.C. PDP8/F computer generated the stimulus that was presented on each trial. A second disk-based D.E.C. PDP8/E computer system determined whether the stimulus would be a signal or a nonsignal and the length of time that a stimulus would be visible, set at 2 seconds unless terminated by a decision response. Following stimulus termination there was a further 2 second interval during which a confidence response could be recorded, and which could also be terminated by a response. Both of the responses were classified by the second computer which also measured the latency of the primary decision response and the value of the confidence response. Summaries of these statistics were printed out after training trials and again after the main session. Following the confidence interval there was always a further interval of approximately 1.5 seconds before the next display appeared. This was processing time required to generate the next display since the lengths of the lines in each was determined randomly prior to every trial according to the signal and non-signal dimensions specified to the PDP8/E prior to the session. The main session of 1000 trials thus could last for a maximum of about 92 minutes.

Observers depressed the left or right of two perspex buttons, lying on either side of a wedge-shaped resting position, to indicate whether a non-signal or signal, respectively, was judged to be present. Confidence was then expressed by depressing one of a line of ten buttons, divided in the middle by a yellow line so that the five on the left were in front of the

nonsignal response button and the five on the right were in front of the signal response button. For both responses the confidence scale ran from "least confidence" at the innermost button lying next to the yellow line, to "most confidence" at the outermost button.

(c) Design

In order to test the generalizability of the previous set of results the design of condition A(i) of Experiment 3 was chosen for replication (see Figure 20). An experimental session then consisted of a block of 50 training trials followed by a main session of 1000 test trials. During training trials signals and nonsignals were presented randomly and with equal likelihood of occurrence. The main session consisted of 20 blocks of 50 trials each, with no gaps between consecutive blocks. Signals were presented randomly within each trial block with a probability of occurrence of .5 during the first four blocks which was lowered to .4, .28, and .16, respectively, during the next three trial blocks and finally held constant at .06 during the final thirteen trial blocks. Due to the continuous nature of the task, and the random fashion in which signals appeared, no abrupt change in signal probability could be perceived.

(d) Procedure

Observers were tested individually and in complete isolation. Watches were removed prior to testing. Initial instructions defined the task as a type of "signal detection experiment" which required observers to distinguish between "signal" and "nonsignal or standard" stimuli. An example of each stimulus was then shown to the observer. The use of the response buttons and confidence scale was then explained, as was the need to make both responses within 4 seconds of stimulus onset and to always use the index

finger of the dominant hand. Observers were told that signals and nonsignals would occur with the same frequency during training trials. Following training the print-out of performance data was checked to see whether the observer was responding correctly. The observer was then told to continue to respond in the same manner during the main session to follow. No further mention was made of relative stimulus frequencies.

(e) Observers

Fourteen females enrolled in the first year Psychology course at the University of Adelaide volunteered to participate in the experiment as part of their course requirements. Ages ranged from 17 to 34 years with a median age of 18 years. All observers had normal or corrected vision.

2. Results

The data of each observer were classified according to three phases of signal probability and five temporal periods for the purpose of statistical analysis. The three phases of signal probability corresponded to the initial phase of constant .5 probability, the phase during which signal probability declined, and the final phase of constant lowered (.06) probability. The first two temporal periods corresponded to the first two phases of signal probability, and the final three were obtained by dividing the final phase of lowered signal probability into three roughly equal periods. This latter classification allowed tests for performance change over the final phase of constant signal probability.

(a) Response probability

(i) Conditional signal response probabilities. Table 11 gives the results of analyses of variance on response probabilities with signal probability as the main factor. The significant effect on the probability of correctly detecting a signal is apparent in Figure 31: an increase in detection probability with the decline in signal probability. There was no additional change in the detection measure over the final three time periods ($F = .280$, d.f. = $2/26$, $p = .75$). The significant effect of signal probability on the proportion of incorrect signal reports, or false alarms, can also be seen in Figure 31. Again the increase in the probability of correctly reporting a signal in its presence was accompanied by an increase in the probability of making a false report in its absence. The decline in false alarm probability over the final

TABLE 11. Results of analyses of variance on response probabilities.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	Correct signal response $P(S/s)$				
	Signal Probability (SP)	2	.303	6.390	.01
	SP x S	26	.047		
(b)	Incorrect signal response $P(S/n)$				
	Signal Probability (SP)	2	.342	3.552	.05
	SP x S	26	.096		
(c)	Signal response $P(S)$				
	Signal Probability (SP)	2	.665	7.038	.005
	SP x S	26	.094		
(d)	Correct response $P(S/s \text{ and } N/n)$				
	Signal Probability (SP)	2	.129	1.421	
	SP x S	26	.091		

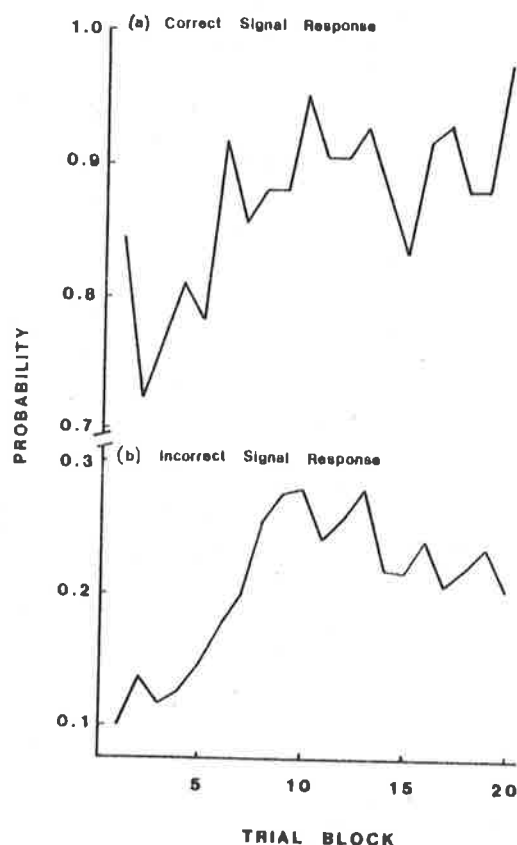


FIGURE 31. Changes in the mean conditional probability of (a) correct and (b) incorrect signal responses over the twenty trial blocks.

three time periods, strongly suggested in Figure 31, failed to reach an acceptable level of significance ($F = 2.635$, $d.f. = 2/26$, $p = .09$).

(ii) Unconditional signal response probability. A strong effect of signal probability on the unconditional probability of making a signal response is reported in Table 11. Figure 32 indicates that signal response probability dropped markedly as signal probability declined. That there was no tendency to increase the proportion of signal responses toward its previous level, as would be necessary for response maintenance to occur, is evidenced by the continued downward trend in signal response probability during the final phase

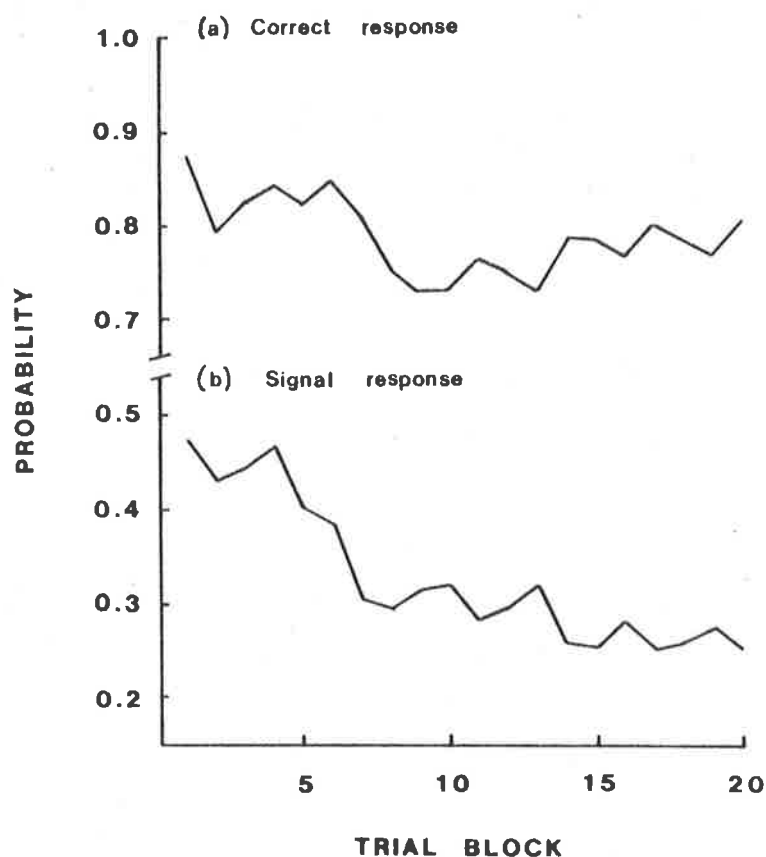


FIGURE 32. Change in the mean unconditional probability of making (a) a correct response and (b) a signal response over the twenty trial blocks.

of constant signal probability. However, as was the case with false alarm probability, an analysis of variance across the final three time periods yielded no significant result ($F = 2.488$, d.f. = 2/26, $p > .10$).

(iii) Correct response probability. The probability of making a correct decision, signal and nonsignal responses combined, was not significantly affected by the change in signal probability (Table 11). It is likely that this was due to an increase in this probability measure, from an initially lowered value, during the final phase of constant signal probability (Figure 32). Such a tendency for any of the performance measures to reverse trends during the final three periods could obscure the main effect of signal probability,

since the data from these periods were combined into one phase of signal probability for the analyses of variance. As a result, a comparison of mean correct response probability between the first phase of signal probability (period 1) and the first period of constant lowered signal probability (period 3) revealed that the probability of a correct response did decline ($F = 6.786$, $d.f. = 1/13$, $p = .02$). The increase over the final three watch periods failed to reach an acceptable level of significance ($F = 2.732$, $d.f. = 2/26$, $p = .08$).

(b) Response latencies

(i) General trends. Owing to the relatively small number of false alarms throughout the task, and even smaller number of missed signals, mainly later in the task, no reliable analyses of variance could be performed on these data. As was the case in Experiment 3, what data were available were sporadic and no consistent trends were evident. The results of analyses of variance for the effect of signal probability on the latencies of the two categories of correct response are given in Table 12. The lack of any significant effect on detection latencies again appears, from Figure 33, to be due to a tendency for the initial trend in this measure to reverse direction, and hence decrease during the final phase of the task, so that the mean level of performance was not greatly different from the levels of the first two phases of the task. This increase over the final three time periods was strong enough to reach significance ($F = 3.672$, $d.f. = 2/26$, $p < .05$). According to the analysis of variance reported in Table 12 signal probability was also not a significant factor influencing the increase in correct rejection latencies apparent in Figure 33. However, when a comparison was made of mean correct rejection latency in the first time period and the fourth time period, the period of highest latency during the final phase of the task, it revealed

TABLE 12. Results of analyses of variance on response latencies and confidence.

(a) <u>Response latency</u> :					
	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(i) Correct signal response					
	Signal Probability (SP)	2	.066	.177	
	SP x S	26	.373		
(ii) Correct nonsignal response					
	Signal Probability (SP)	2	1.535	1.998	
	SP x S	26	.768		
(b) <u>Response confidence</u> :					
	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(i) Correct signal response					
	Signal Probability (SP)	2	1.172	.332	
	SP x S	26	3.529		
(ii) Correct nonsignal response					
	Signal Probability (SP)	2	2.887	.548	
	SP x S	26	5.263		

that a significant increase had occurred between these two points ($F = 5.746$, $d.f. = 1/13$, $p = .03$). There was no significant change in correct rejection latency over the final three time periods ($F = .767$, $d.f. = 2/26$, $p > .75$).

In order to determine whether the initial divergence followed by a convergence of the two reaction times, as suggested in Figure 33, was significant, a difference score was computed for each block of trials on the

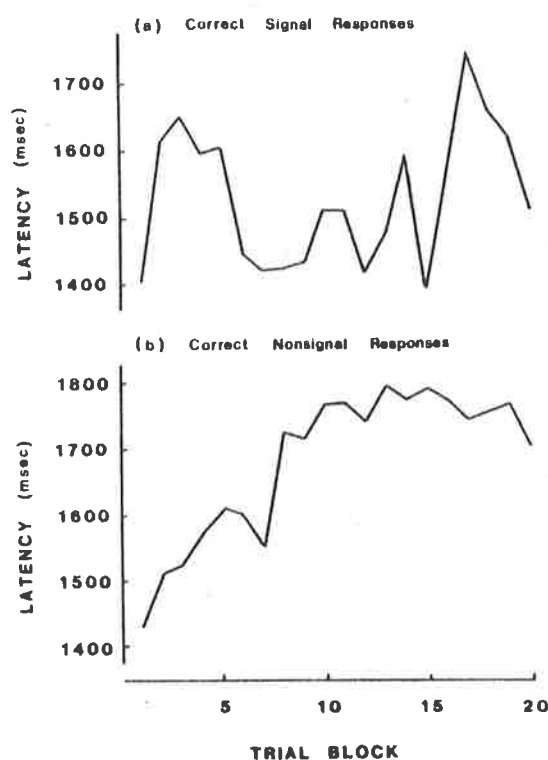


FIGURE 33. Change in the mean latencies of (a) correct signal and (b) correct nonsignal responses over the twenty trial blocks.

data of each observer by subtracting the mean latency of correct rejections from the mean latency of correct detections. This difference score was then subjected to an analysis of variance and showed significant changes over the first three time periods ($F = 10.894$, d.f. = 2/26, $p < .0005$) and the final three time periods ($F = 4.915$, d.f. = 2/26, $p < .05$). That is, the latencies of detections declined relative to those of correct rejections over that part of the session during which signal probability declined, and increased relative to correct rejection latency over the final phase of lowered, but constant, signal probability.

(ii) Differences among the four response categories. Table 13 contains the mean response latency for each of the four decision categories. Paired-sample t-tests were performed to test for mean differences. For both signals and nonsignals correct reports were made faster than incorrect reports ($t = -18.84$, d.f. = 246, $p < .001$, and $t = -8.74$, d.f. = 126, $p < .001$, respectively), and while there was no significant difference between the two types of incorrect report ($t = -1.75$, d.f. = 97, $p > .08$) signal detections were made faster than correct rejections ($t = -5.41$, d.f. = 279, $p < .001$), even though, as indicated in Figure 33, they were slower initially.

TABLE 13. Mean latencies and confidence for the four categories of response.

(a) <u>Response latency</u> (msec)			
<u>Correct signal</u>	<u>Incorrect signal</u>	<u>Correct nonsignal</u>	<u>Incorrect nonsignal</u>
(S/s)	(S/n)	(N/n)	(N/s)
1529	2015	1681	1870
(b) <u>Response confidence</u> (5-category scale)			
3.23	2.04	3.13	2.79

(c) Response confidence

(i) General trends. As with response latencies no reliable analysis could be performed on the confidence expressed in false alarms and false rejections. The results of analyses of variance for the effect of signal probability on confidence in the two types of correct response are summarized in Table 12. As reported in the table there was no significant effect on the confidence expressed in either response. There was also no significant change over the final three time periods ($F = .969$, $d.f. = 2/26$, $p > .35$ for detections; $F = 1.401$, $d.f. = 2/26$, $p > .25$ for correct rejections). Comparisons of mean confidence in the initial period of the task with mean confidence in periods during the final phase yielded only one result worthy of mention. This was a decrease in correct rejection confidence between the first and fifth time period, as can be seen in Figure 34, but which failed to reach an acceptable level of significance ($F = 3.277$, $d.f. = 1/13$, $p = .09$).

By use of the procedure employed with response latencies a difference score was computed between the mean confidence expressed in detections and correct rejections. An analysis of this measure confirmed that the increase in detection confidence relative to correct rejection confidence from the second to the fourth time period, as suggested in Figure 34, was significant ($F = 4.740$, $d.f. = 2/26$, $p < .05$). The subsequent convergence of the measures over the final two time periods was not significant.

(ii) Differences among the four response categories. The mean confidence for each of the four decision categories is given in Table 13. Paired-sample t-tests revealed that both types of correct response resulted in more confidence than their incorrect counterpart ($t = 23.12$, $d.f. = 244$, $p < .001$ for signal responses; $t = 8.44$, $d.f. = 123$, $p < .001$ for nonsignal responses) and that although there was no significant difference between

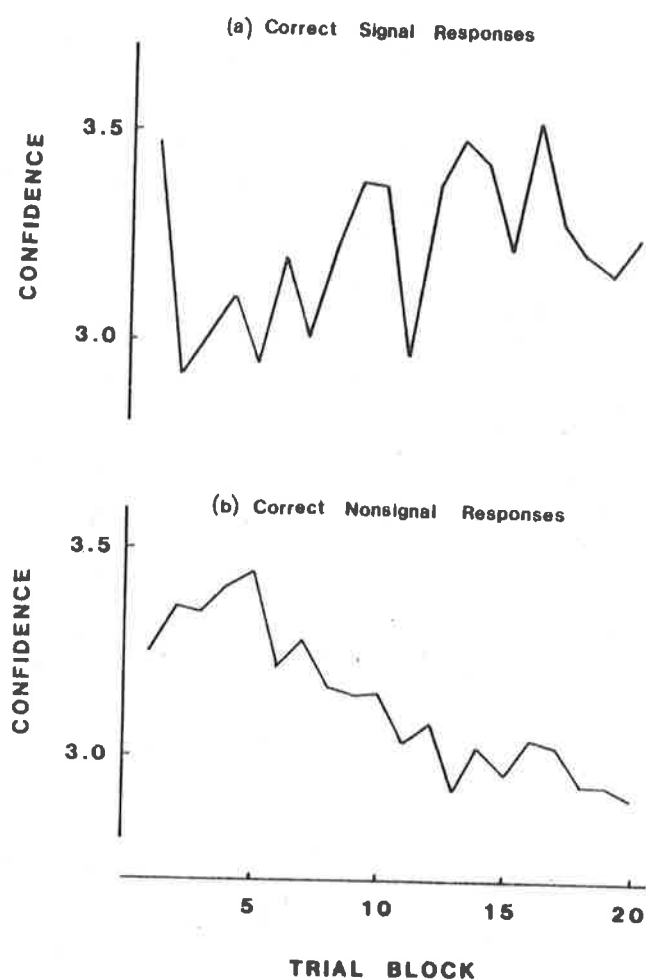


FIGURE 34. Changes in the mean confidence expressed in (a) correct signal and (b) correct nonsignal responses over the twenty trial blocks.

the confidence means of the two types of correct report ($t = 1.63$, d.f. = 279, $p > .10$), false alarms were made less confidently than false rejections ($t = 4.31$, d.f. = 94, $p < .001$).

(d) Signal detection measures and theoretical decision parameters

(i) Effect of signal probability. Although the number of misses and false alarms were too few to permit an analysis of their latency and confidence, enough were made to calculate values of d' and β for 11 of the 14 observers at each of the three levels of signal probability. The results of analyses of variance on these measures is given in Table 14. The nonsignificant effect of signal probability on d' combined with its significant effect on β indicates that the changes in performance observed as signal probability declined were the result of a shift in response bias rather than the ability to detect signals. As can be seen in Figure 35, a decline in β accompanied the fall in signal probability, as opposed to the increase that would be demanded by the optimal values of β specified by signal detection theory.

TABLE 14. Results of analyses of variance on signal detection measures.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	Sensitivity (d')				
	Signal Probability (SP)	2	.010	.102	
	SP x S	20	.099		
(b)	Bias (β)				
	Signal Probability (SP)	2	12.793	3.589	.05
	SP x S	20	3.563		

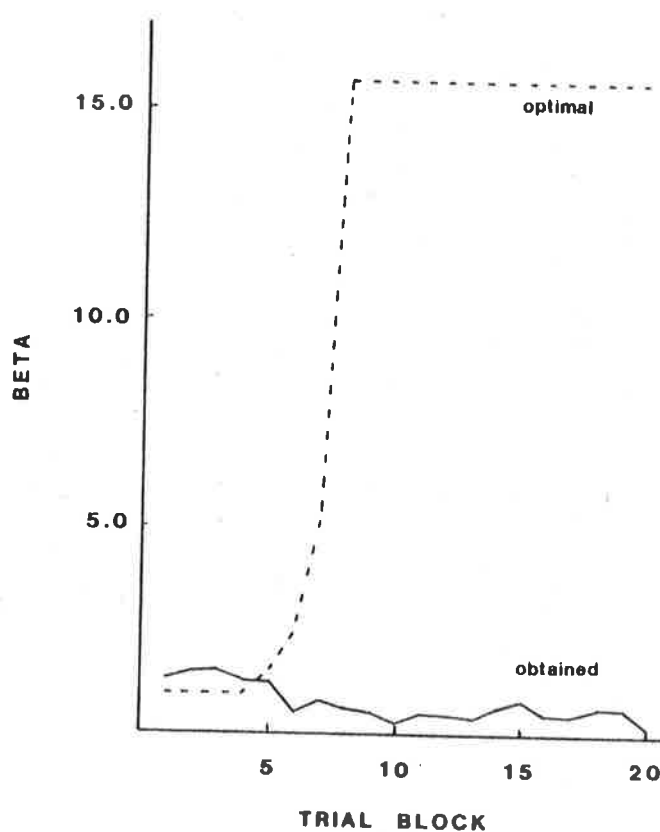


FIGURE 35. Changes in the observed group values of β compared with the optimal values specified by signal detection theory over the twenty trial blocks.

(ii) Relationship between β and $(1-c)$. Figure 36 compares changes in β , calculated from group mean conditional signal response probabilities, and $(1-c)$, calculated from group mean unconditional signal response probabilities, over each of the twenty trial blocks. Practice trials were used in addition to calculate the value of c in each trial block. Indicative of the operation of some form of a stabilization process, β did not change considerably until after the fifth trial block, in which there occurred the first large decline below zero in $(1-c)$ due to the drop in local signal probability. The subsequent decline in β continued beyond the point at which $(1-c)$ reached its maximum negative value, and might again be interpreted to be the result of the interval necessary to raise subjective estimates of local and cumulatively-based signal

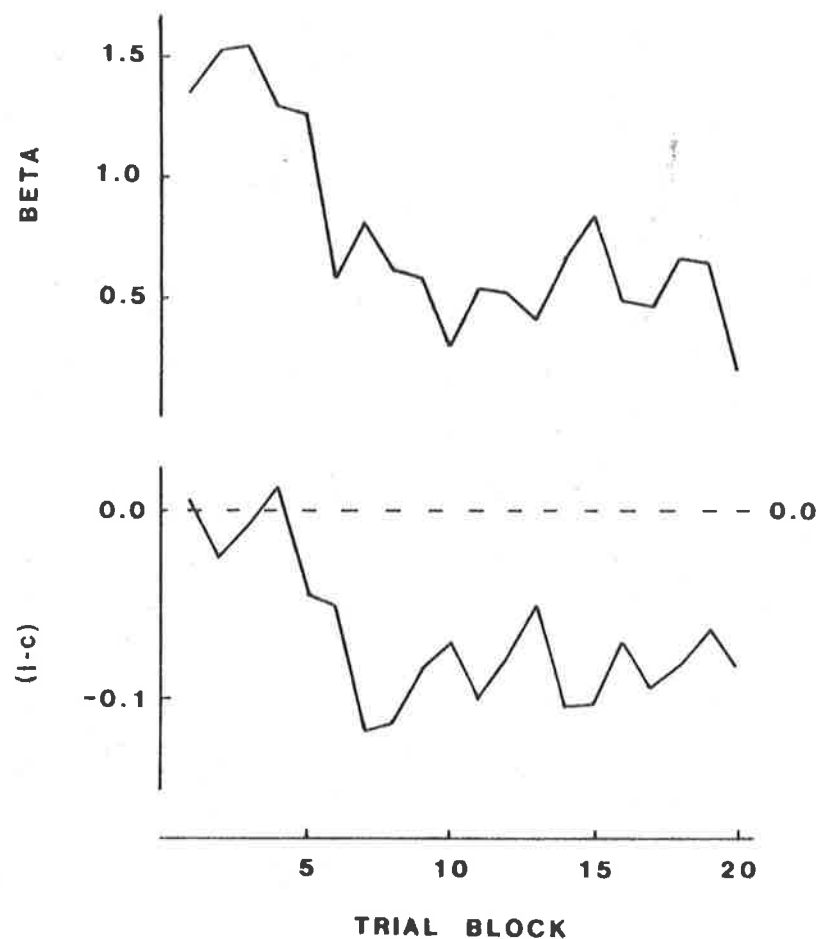


FIGURE 36. A comparison of changes in group values of β and mean values of $(1-c)$ over the twenty trial blocks.

probability and then act upon any discrepancy detected between the two.

As a consequence, the Pearson product-moment correlation coefficient between group values of β and mean values of $(1-c)$ in each of the twenty trial blocks again confirmed that there was a high positive relationship ($r = .689$; $n = 20$; $p < .0005$, 1-tailed) which was further enhanced when the mean value of $(1-c)$ in each trial block was compared with the group value of β in the subsequent trial block ($r = .728$; $n = 19$; $p < .0005$, 1-tailed).

In Figure 36 there is no indication of an unbroken trend for β to increase after signal probability was held constant, as would be predicted on the basis of a merger between local and cumulatively-based signal probability. However,

there is also no continuous tendency for $(1-c)$ itself to return toward zero. That is, the difference between the two response probability measures was maintained at a negative value later in the task. After an initial movement upwards toward zero from the seventh to the thirteenth trial block, $(1-c)$ declined and then appeared to oscillate around a value somewhat less than zero. In a similar manner, β tended to increase from the tenth to the fifteenth trial block after which it showed no further upward movement, a downward trend in fact being indicated. Hence it would appear that a fairly large discrepancy was being tolerated between the local and cumulatively-based probability of a signal response.

3. Discussion

(a) Correspondence to previous results

(i) Relationships among performance measures. The pattern of the relationships shown among the probability, latency and confidence measures of performance in the discrimination tasks has almost entirely been reproduced in the detection task, and for the most part there is agreement with the distance-from-criterion hypothesis in that form presented in Chapter 1 (section B.3.(d)). During that part of the task in which the conditional probabilities of a signal response increased, indicating a decrease in the criterion for a signal response, there were increases in both the speed and confidence of signal reports relative to nonsignal reports. Consistent with the obtained group value of β of less than 1.0 throughout most of the task, correct signal reports were made more rapidly overall than correct nonsignal reports. No overall difference was found between the confidence expressed in the two types of correct response. This was probably due to the fact that although the confidence expressed in correct signal reports decreased

during the task while that of correct nonsignal reports increased, it was initially at a lower level (see Figure 34). Finally, for both signal and nonsignal reports, correct responses were made more rapidly and confidently than incorrect responses.

The only evidence opposed to the distance-from-criterion hypothesis is contained in the finding that false rejections of a signal were made more confidently than false alarms. As the hypothesis assumes degree of confidence to be a direct function of the distance an observation lies from the decision criterion, this implies a criterion to the right of the crossover point of the signal and nonsignal distributions, and hence a value of β greater than 1.0. In fact, 53 percent of those false rejections which could be paired with false alarms for mean comparisons were made during the first five blocks of the task where the group value of β was always greater than 1.0. However, since β declined and remained below 1.0 after the fifth trial block, confidence in false rejections should then have decreased and false alarms increased, and no overall difference been found between the confidence of the two types of report. The only way that observations giving rise to false rejections could still have been distributed further from the decision criterion over the final fifteen blocks, and so help produce the overall results, is if the decline in signal probability led to an increase in the variance of the signal distribution relative to the nonsignal distribution. This was the only rationale by which the hypothesis could explain the finding of longer false alarm than false rejection reaction times in Experiment 1. Again this should have been detected as a decrease in d' as signal probability declined, whereas this was not found. In fact, d' increased slightly during the task from 1.97 in the first time period to 2.16 in the final period.

(ii) Main effect of level of signal probability. While there was little evidence against the distance-from-criterion hypothesis, the response optimization model of signal detection was shown again to be clearly inappropriate in a situation of time-varying signal probability. Rather than an increase in β , as would be demanded by the optimal values specified by signal detection theory, there was a significant decline in β to coincide with the decline in signal probability. Also consistent with this implied increase in bias toward making a signal report were increases in both the speed and confidence of correct signal reports relative to correct nonsignal reports during the portion of the task in which signal probability declined.

Although the decline in β is consistent with both the response maintenance hypothesis of Williges (1969) and the response stabilization model, the trends observed in other response measures again give clear evidence against response maintenance. The change in β was not brought about by a highly disproportionate increase in false alarm rate relative to detection rate (Figure 31), as would be necessary for response maintenance to take place. Hence there was no change in the sensitivity measure, d' , as signal probability declined, and the results reinforce the findings of the discrimination tasks to lend further support to the response stabilization model.

(iii) Performance across the final phase of constant signal probability.

The prediction of the response stabilization model that performance over the entire final phase of constant low signal probability would indicate a decrease in the amount of bias toward making a signal response was based on the assumption that $(1-c)$ and, therefore, β would follow the difference between the objective local and cumulatively-based probability of signal occurrence. This difference increased toward zero from its maximum negative value. It may be recalled, in the discrimination task which corresponded

to the present detection task in its design, $(1-c)$ showed a fairly continuous upward trend during this period, eventually reaching zero, and this was accompanied by an upward trend in β . A strong positive relationship again was shown between $(1-c)$ and β through the present task as a whole. However, the tendency for $(1-c)$ to return to zero did not continue throughout the final phase of the task. Instead, the initial upward trend in $(1-c)$ was followed by a tendency to oscillate around a value somewhat less than zero. Similarly, β at first showed a tendency to increase, this tendency ceasing, and even reversing, after $(1-c)$ showed no further upward movement. Hence, the relationship between $(1-c)$ and β indicates that a relatively large discrepancy was being tolerated between local and cumulatively-based signal response probability.

Despite the fact that a larger value of $(1-c)$ was tolerated, and the upward trend in β was not continuous, change over the final phase of the task in other performance measures indicate that response bias was being altered in a manner consistent with the stabilization model. Although not significant, unless a liberal level of .10 is accepted, the decrease in false alarm probability suggests that there was a decrease in the bias toward making a signal response. This probability is a direct measure of the proportion of the nonsignal distribution lying to the right of the decision criterion, and its decrease indicates that the criterion has moved to the right from its previous position.

The strongest evidence for such an increase in β is provided by the significant increase found in the reaction times of correct signal responses. This increase was not due to some nonspecific increase in reaction times resulting from the amount of time spent at the task, as this would have led to an increase also in the latency of correct nonsignal responses. Rather,

it may be that the reaction time measure was more sensitive to the effects of response bias than the probability and confidence measures. That it is more sensitive to experimental manipulation than associated measures of performance has been demonstrated in vigilance studies (reviewed by Buck, 1966). According to a distance-from-criterion interpretation of response latencies, increased detection latencies imply that observations giving rise to correct signal reports now lie closer to the decision criterion. Since, as mentioned previously, d' tended to increase during the task, it is not likely that this took place through a leftward movement of the signal distribution but, instead, through a rightward movement of the criterion and so reflects at least the initial increase in β .

Besides the distance-from-criterion interpretation, an increase in detection latency indicates an increase in the criterion for a signal response according to models of the detection process in which the observer is assumed to make a series of observations of the sensory evidence that is present during each interval, and which can vary from moment to moment within that interval. Two such models, as presented in their earliest and least developed form, are the 'accumulator' model of Vickers (Vickers, 1970), and the 'multiple observations' model of Pike (1973). Both models propose that on each trial the observer makes repeated observations at a steady rate, each observation providing evidence that either a signal or nonsignal is present. A decision is reached and response initiated when either a critical (criterion) number of observations have been made in favour of one of the two response alternatives, according to Pike's model, or the evidence accumulated during the series of observations has reached the critical magnitude required for either of the responses, according to Vickers' model. (Both of these will be discussed in more detail in Chapter 6). Since each observation takes a finite period

of time, reaction time to a signal of constant magnitude is directly proportional to the value of the criterion for a signal response. Hence an increase in detection latency would indicate that the criterion, in terms of either a critical number of observations or a critical total of sensory evidence, has been increased.

A decrease in the bias toward a signal response is therefore indicated by the significant increase in detection latencies according to differing conceptions of the detection process. Since a decrease in bias is also suggested by the marginal effect on the probability of a false report and an initial tendency for the signal detection measure, β , to increase, those trends that were observed over the final phase of the task are in general agreement with the response stabilization model. Taken in conjunction with the results of the discrimination tasks, it thus appears that observers do adjust the value of their decision criterion in order to reduce any excessive discrepancy between the local and cumulatively-based probability of a signal report, but that the amount of discrepancy which is tolerated may depend upon the type of task employed.

(b) The control of ongoing performance in vigilance

The applicability of the response stabilization model to the field of vigilance has been demonstrated by the general concordance between the results of the discrimination tasks and those of the detection task more characteristic of the traditional vigilance study. The conclusion is then that at any point during a vigilance task the observer may alter the degree of bias he has toward making a signal response if the present proportion of signal responses either exceeds or fails to reach some expected proportion. Further, the observer may switch between different strategies in order to reduce the imbalance, perhaps adopting a simpler strategy when it presently

appears to achieve the desired result. This was shown in condition B of the discrimination tasks when observers seemed to abandon the more complex decision routine specified by the stabilization model and adopted a strategy of altering their response bias so as to simply produce responses in accordance with their long-term probability of occurrence. However, the response stabilization goal of reducing the discrepancy between that proportion and the present proportion was still satisfied.

The literature on vigilance has recently been cited by Rabbitt (1979) to exemplify the general need for a transition to models of behaviour in which observers exert active control over their ongoing performance. Such models are seen to be necessary because in vigilance, as well as other types of experimental task, observers "use their acquired information about the structure of particular tasks, and about the external world in order to actively control their momentary perceptual selectivity and their choice of responses" (Rabbitt, 1979, p. 129). Such an assumption is intrinsic to the response stabilization model in that the model is concerned with the manner in which the observer utilizes stored parameters calculated from previous task experience in order to determine whether the present decision criterion should be set at some new value. The notion that the observer uses feedback from the task in order to adjust his decision criterion is also explicitly stated in the model of Indlin (1976) and will be discussed in the next chapter.

CHAPTER 4.

A. ADJUSTMENT OF THE DECISION CRITERION THROUGH FEEDBACK

1. Indlin's (1976) Response Regulation Model

(a) Main features of the model

The notion that the observer adjusts his decision criterion on the basis of information drawn from his previous response sequence is fundamental to a model of the human observer proposed by Indlin (1976). This model will be termed response regulation since the observer is considered to behave according to the principles of a specific type of feedback system, a regulator. The purpose of such a system is to maintain some aspect of its own performance (output) at a desired, or reference, level (input) as shown in Figure 37. At any moment during its course of operation, the value of the output of the system serves as a feedback signal which is compared with the reference signal, an input which specifies the desired system output. Whenever there is a discrepancy between the two an error signal is produced which leads to corrective action being taken by the system in order to restore its output to the desired value.

According to Indlin (1976) the observer alters his decision criterion whenever there is a discrepancy between the recent proportion of a certain response and the long-term proportion of that response. Specifically, in a detection task the feedback signal is given by the recent proportion of signal responses which is

$$p(S) = p(s)p(S/s) + p(n)p(S/n) \quad (4)$$

where $p(S)$ is the unconditional probability of a signal report, $p(s)$ and $p(n)$ are the a priori probabilities of signal and noise, respectively, and $p(S/s)$ and $p(S/n)$ are the conditional probabilities of a correct detection and false

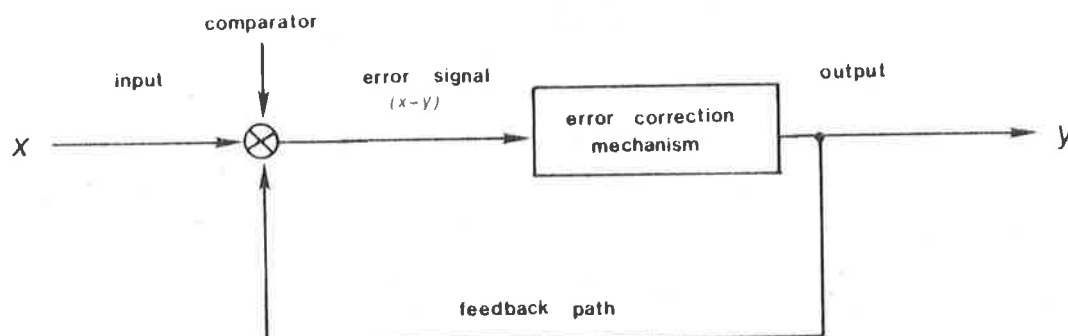


FIGURE 37 Diagram of a simple feedback system in which the present output (y) of the system is compared with the input (x) which has been applied to it. At the comparator the value of the output is subtracted from that of the input and whenever the result differs from zero an error signal is produced which activates the error correction mechanism.

report, respectively. The reference signal, or desired proportion of signal responses, is the proportion of signal responses made over the entire event sequence. In steady-state conditions, that is when signal probability is constant and performance has stabilized, the reference signal is considered to be the a priori probability of signal occurrence, $p(s)$. Whenever signal response probability differs from a priori signal probability, an error signal is produced which requires the decision criterion to be altered in such a way that response probability will again tend to match objective probability. Hence, the goal of the observer is to respond in a manner that will satisfy the equation

$$p(s) = p(s)p(S/s) + p(n)p(S/n) \quad (5)$$

The latter equation can be rearranged to yield a measure of β which is based solely on observed conditional response probabilities. Assuming that the payoff matrix is symmetrical, the optimal value of β is given by the ratio, $p(n)/p(s)$, and solving equation 5 for this ratio gives

$$\frac{p(n)}{p(s)} = \frac{1-p(S/s)}{p(S/n)} \quad (6)$$

Indlin argued that values of β based on this measure would show a closer correspondence to optimal values than those obtained through the usual calculation of β . He showed that a plot of the right-hand portion of the last equation against $p(n)/p(s)$ closely approximated a straight line with unit slope, as opposed to a line with a slope of less than unity as is typically obtained when measured β is plotted against $p(n)/p(s)$ (Green & Swets, 1966). Thus equation 6 reduces the "conservatism" usually found in measures of β , overly-high values when optimal β is low and overly-low values when the optimum is high. In fact, any pair of detection and false alarm probabilities that satisfy the decision goal specified by equation 5 will yield a value of β in equation 6 that is numerically equivalent to the optimal value specified by signal detection theory. This necessarily follows from the fact that equation 6 is a rearrangement of equation 5.

(b) Response regulation as probability-matching

Due to its particular choice of feedback and reference stimuli, the response regulation model can be thought of as providing a feedback principle by which "probability-matching" will take place (Chapter 2, section B.3.(c)(iii)). Probability-matching was proposed by Parks (1966) in his signal detection model of recognition memory, according to which subjects in a recognition task adopt a cutoff position of the decision criterion such that the proportion of items that will be reported as having been previously presented will be given by

$$p(S) = kp(s) \quad (7)$$

where $p(s)$ is here the a priori probability that a previous item will be presented and k is a constant determined by the payoff matrix. In a mathematical evaluation of the probability-matching hypothesis, Thomas and Legge (1970) assumed that k takes on a value of unity when the payoff matrix

is symmetrical, and in this instance the decision rule given by equation 7 is identical to the observer's goal given by equation 5. Further, Thomas and Legge (1970) showed that a mathematical prediction of the probability-matching hypothesis is that obtained values of β will be conservative relative to the optimal values that are expected according to signal detection theory. Equation 6 thus presents a relationship in which obtained values of β will be non-conservative since they are based on conditional response probabilities that are assumed to be continually regulated according to a probability-matching process.

2. Response Regulation Versus a Response Stabilization Process

(a) Response to change in signal probability

Whereas the response regulation model is mainly concerned with rather minute adjustments of the decision criterion that could be due to variation in the recent local probability of a signal response, the response stabilization model attempts to explain the manner in which the observer copes with change in both long- and short-term signal response probability. According to the response stabilization model, whenever there is a discrepancy between local and cumulatively-based signal response probability, the particular adjustment that is made to the decision criterion will depend not only upon the present difference between the two response probabilities, but also upon whether there has been another recently noted discrepancy (see Figure 16). An important point to note here is that minor fluctuations in the proportion of signal responses which do not appreciably affect long-term, or cumulatively-based, signal response probability will be responded to in the same way by both models.

A difficulty for the response regulation model is encountered when the change in local response probability is due to a change in a priori signal

probability, so that long-term signal response probability will ultimately be affected. According to Indlin's response regulation model, the observer maintains signal response probability at his estimate of the level of a constant a priori signal probability. Thus, in a strict sense, the model might be deemed inappropriate in a situation where a priori signal probability varies over time. However, Indlin (1976) has stated that, under such conditions, the observer is expected to find a new position of the decision criterion "by trial and error" that will satisfy the goal specified by equation 5. That is, he should adopt a reference signal given by the value to which a priori signal probability has shifted.

This could occur when observers are aware that a particular change in signal probability will take place, or are given external feedback about their performance. However, in the absence of accurate information being provided to the observer, it must be assumed that the reference signal, defined as the proportion of signal responses during all preceding experimental trials, will take on a value corresponding to cumulatively-based signal probability. Response regulation then predicts that there will be either perfect or cumulative probability-matching. In either case, the direction of movement that will be observed in the observer's decision criterion can be predicted from equation 6.

For instance, a gradual decline in signal probability will result in an increase in the value of the ratio on the left of the equation, irrespective of whether the estimate of objective signal probability is based on the present local probability of signal occurrence, or its probability calculated over the entire sequence of trials. Hence an increase is demanded in the value of the observer's criterion as measured by the right-hand portion of the equation. Assuming that the observer makes orderly adjustments of the decision criterion

rather than employing some guessing procedure, this will require an increase in the complement of the probability of a correct detection, and a decrease in that of a false report. These changes are equivalent to an increase in β , and so reflect a criterion that is tending toward the optimal value as specified by the ideal observer hypothesis. Such a tendency was not found in any of the previous experiments, none of which gave any forewarning to observers that a particular change in signal probability would occur.

(b) A possible hierarchy of feedback systems

The possibility exists that observers behave in the manner suggested by response regulation under steady-state conditions where a priori signal probability remains constant over the entire experimental period, or under conditions of varying signal probability when they are instructed about the particular response ratios that should be adopted. However, any change in the rate of making signal responses during a task where the response ratio varies under conditions of imperfect prior information could be due to a real change in signal probability, or to an inappropriate degree of bias toward making a signal response. Thus, under such conditions, response stabilization might be a higher-order process, which serves the necessary function of creating a balance between the local proportion of signal responses and the long-term expected proportion. If this is so, a change in cumulatively-based signal response probability (the input to the response regulation system) might be a trigger for the response stabilization mechanism to intervene. When its goal has been achieved, it will cease to function until there is another change in long-term signal response probability. Meanwhile, the regulator will operate unimpeded with its input 're-set' to the new value of cumulatively-based signal response probability.

It may be argued that such a hierarchy of feedback control systems

assumes an unnecessarily high level of complexity, especially since response stabilization has been shown to make essentially the same predictions as response regulation under conditions of constant a priori signal probability, and gives a better account of the effects of a time-varying signal probability. However, it might provide an explanation for the criterion adjustments found among observers in condition B of Experiment 3. It may be recalled that these observers initially made those criterion adjustments specified by the response stabilization model, but, later in the task, appeared to adjust their decision criterion so as to simply produce a response rate which matched cumulatively-based response probability.

B. EXPERIMENT FIVE : DECISION CRITERION ADJUSTMENTS DURING A PROGRESSIVE DECLINE IN SIGNAL DETECTABILITY

The preceding discussion would seem to imply that it may be impossible in many instances to conclude whether a response regulation or response stabilization process, or a combination of the two, gives the better account of shifts in the observer's decision criterion. This is especially true, since the response regulation model is concerned with performance during long periods of constant a priori signal probability. Changes in criterion which have been observed when signal probability varies, and which are in the opposite direction to the predictions of response regulation, must be explained either as the consequence of the observer behaving according to the goals of some other process, or as the inability of the mechanism to function under such dynamic conditions. If the general assumption is maintained that the position of the decision criterion is controlled through some ongoing feedback process, then either explanation requires the specification of some second

process akin to response stabilization. If this additional process can be shown to give the better account of decision behaviour in a situation which meets the necessary requirements for response regulation to occur, and further can account for performance in situations where response regulation is deemed to be inapplicable, then the response regulation model must be rejected and the other model considered as more comprehensive.

In order to test between the two models, a situation must be contrived which meets the full requirements for the operation of response regulation, but might also call for the intervention of response stabilization. Such a situation will arise when a priori signal probability is held constant over a series of trials, but signal response probability can be predicted to vary due to a progressive decline in signal detectability. The outcome of such a manipulation will depend upon whether the observer uses feedback to determine an appropriate position of his decision criterion at each level of detectability and, if so, which particular feedback process is employed. The predictions of an open-loop decision model which does not operate on feedback, but which specifies a constant decision rule, will now be contrasted with the predictions of the response regulation and stabilization feedback models.

1. Predictions of Three Models

The experimental situation to be considered is one in which a priori signal probability is held constant at 0.5 throughout, but the detectability of signals is progressively lowered. At no time is the observer given information that signal detectability is about to change, or advised of the present level of detectability. In order to present the predicted trends in performance associated with each model, the distance-from-criterion interpretation of response latencies and confidence will be assumed, as will be symmetry of

the payoff matrix. In addition, it will be assumed that the observer initially places his decision criterion at the intersection of the signal and nonsignal distributions so that β takes on the appropriate "equal-odds" value of 1.0.

(a) Fixed sensory cutoff model

The first model is one which proposes that the observer does not adjust his decision criterion on the basis of feedback about changes in the ratio of signal to nonsignal responses, but instead keeps it fixed in order to satisfy some constant decision rule. Green and Swets (1966) have shown that several such rules are optimal if the observer employs a likelihood ratio decision criterion, and the one to be considered first is that the observer attempts to satisfy the Neyman-Pearson objective, that is, maximize the probability of a correct detection while holding the probability of a false alarm constant. Evidence that this rule might be obeyed is provided by Curry, Nagel and Gai (1977) who randomly presented observers with stimuli of varying detectability during a discrimination task. The majority of observers showed a tendency for correct detection probability to increase, but false alarm probability to remain almost constant, as detectability increased.

In the present situation this will occur if observers maintain their decision criterion at a fixed position along the sensory continuum when signal detectability declines as shown in Figure 38. That is, the same amount of sensory evidence will be required on each trial in order to report a signal, irrespective of the detectability of the signal and hence its position along the sensory continuum. This model is appealing in its simplicity as it does not need to suppose that the observer converts sensory activity into a measure of likelihood ratio. Indeed, such a transformation would not be accurate unless observers are given information about the prevailing level of detectability

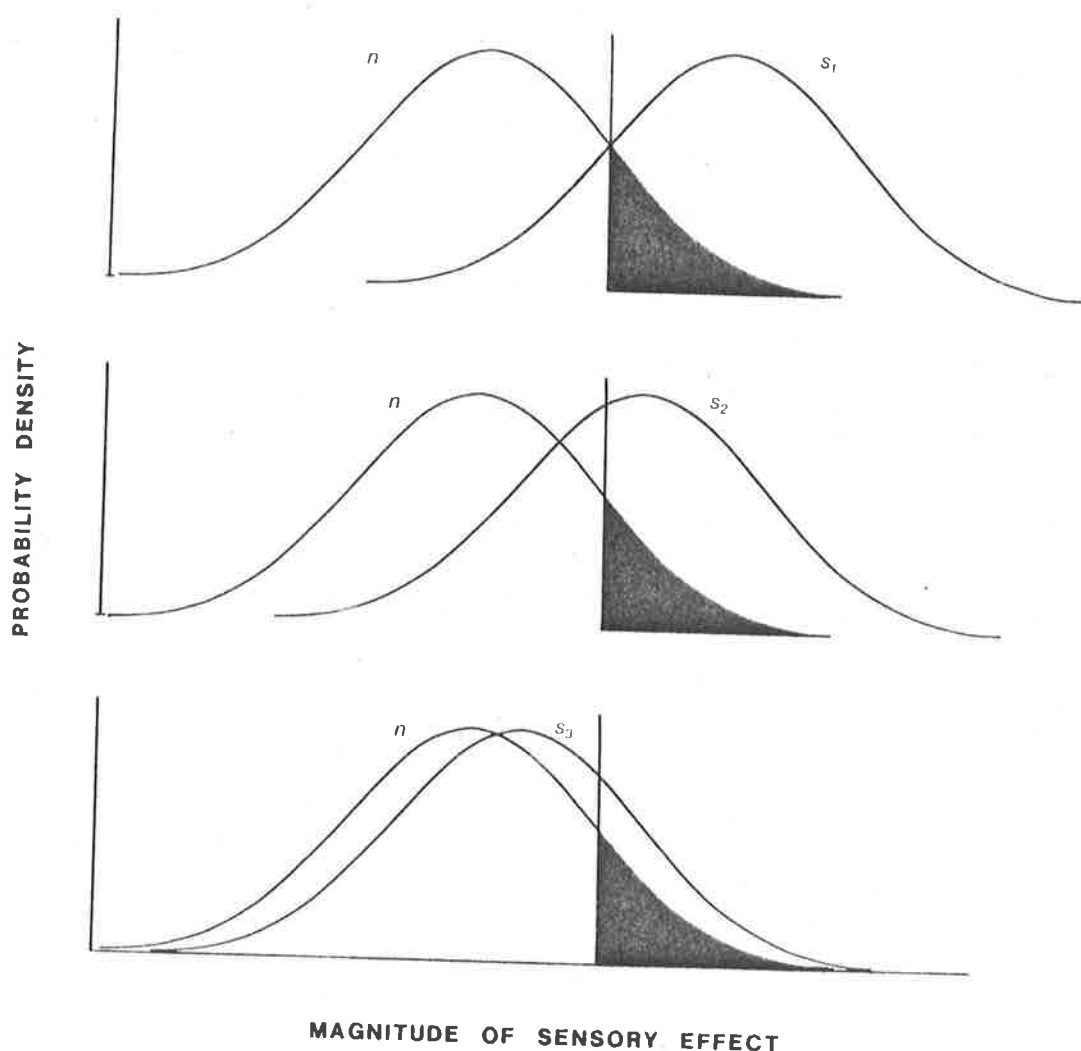


FIGURE 38. The fixed sensory cutoff model: the observer maintains a fixed position of the decision criterion as the signal distribution declines from s_1 to s_3 . The probability of a false alarm (shaded portion of n) remains constant as the unconditional probability of a signal report declines.

so that each sensory observation is converted into the likelihood that it arose from the distribution from which signals are currently being sampled, and not one of the other possible distributions.

The patterns of performance associated with a fixed sensory cutoff are given in Figure 39. As detectability declines, the signal distribution moves leftward past the cutoff position so that the proportion of correct

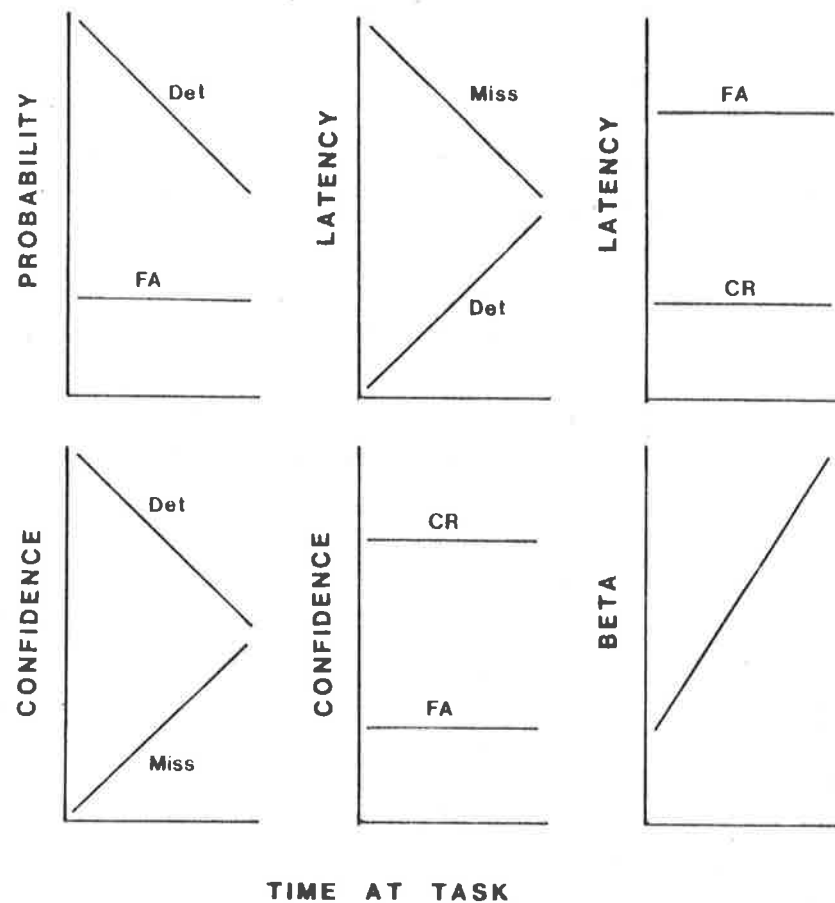


FIGURE 39. Patterns of change in performance that are expected to follow from a downward movement of the signal distribution past a fixed sensory cutoff, as shown in figure 38. The symbols used are Det (correct detections), FA (false alarms), CR (correct rejections) and Miss (missed signals).

signal responses will decrease and that of missed signals will increase. Since observations giving rise to correct signal reports will then lie progressively closer to the decision criterion there will be an increase in detection latencies and decrease in detection confidence. At the same time, observations giving rise to missed signal reports will lie progressively further from the criterion, and so result in decreased missed signal latencies and increased missed signal confidence. Since the position of the decision criterion along the nonsignal distribution does not vary, there will be no effect on the proportion of correct rejections and false alarms, and therefore no change

in the latency and confidence of these responses. Since the distance between the means of the signal and nonsignal distributions will continually decrease, there will be a decline in the sensitivity measure, d' . Meanwhile, since the ordinate value of the decision criterion on the signal distribution increases while it remains constant on the nonsignal distribution, this will be accompanied by an increase in β .

(b) Response regulation

(i) In order to respond according to a priori signal probability. A

second possibility is that the observer adjusts the position of his decision cutoff according to a response regulation process, that is, in order to maintain local response probability at the level of a priori probability. In this case, a decrease in the proportion of signal reports, due to a smaller area of the signal distribution lying to the right of the decision criterion, will demand that the decision criterion be shifted to the left in order to restore the proportion of signal reports to its previous value. This corrective action will be continually triggered as detectability declines and the signal distribution moves leftward. If the observer initially adopts an "equal-odds" criterion, then the regulation process will result in a decision criterion that moves continually to the left in order to retain its position at the intersection of the signal and nonsignal distributions, as shown in Figure 40.

The patterns of performance predicted by a response regulation process are shown in Figure 41. Since the position of the decision criterion should be maintained at the intersection of the signal and nonsignal distributions, the movement to the left of the signal distribution will result in a decrease in the proportion of both types of correct response (detections and correct rejections) and an increase in the proportion of both types of incorrect response (false alarms and missed signals). Since observations which give

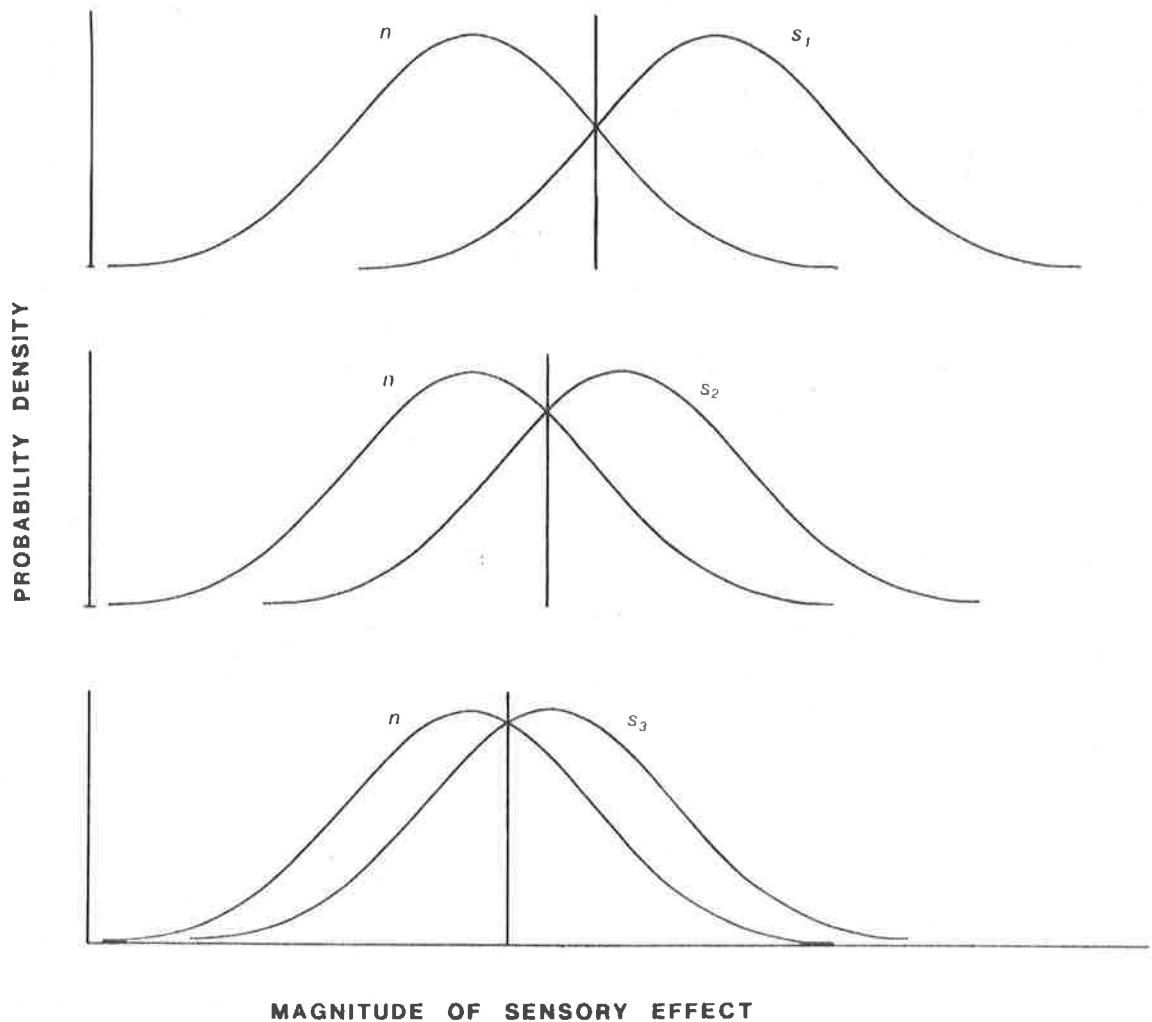


FIGURE 40 The response regulation model: the observer maintains the decision criterion at the intersection of the s and n distributions in order to maintain at a constant value the proportion of observations lying to the right of the criterion.

rise to correct responses will lie progressively closer to the decision criterion, the latencies of detections and correct rejections will increase, while their confidence will decrease. At the same time, observations giving rise to incorrect responses will lie progressively further from the decision criterion so that the latencies of false alarms and missed signals will decrease while their confidence will increase. As the distance between the means of the

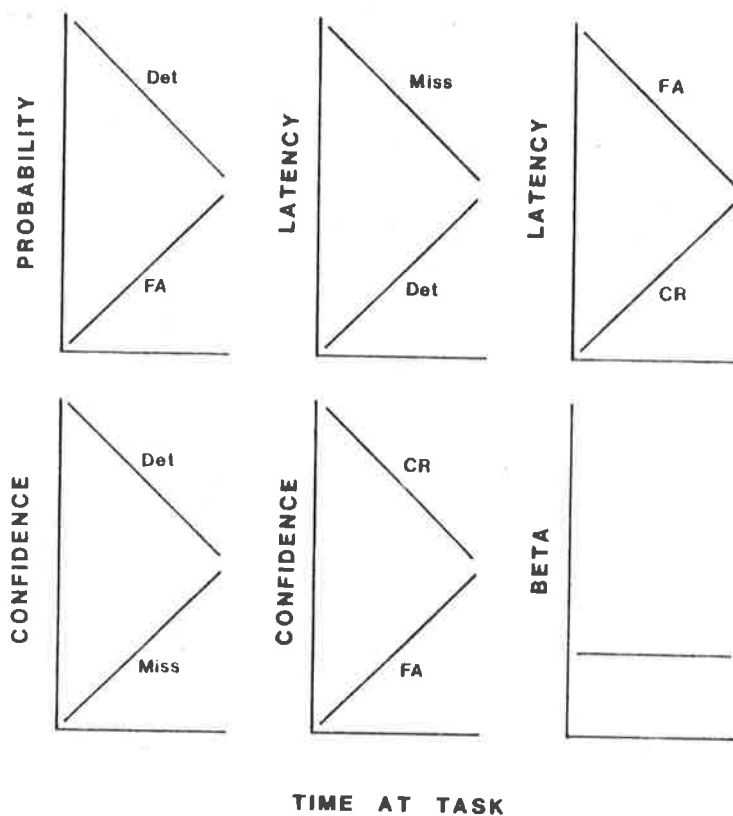


FIGURE 41. Patterns of change in performance that are expected to follow from the response regulation process shown in figure 40. The symbols are the same as those used in figure 39.

signal and nonsignal distributions decreases, there will be a decline in measured sensitivity, d' . However, the value of β will be unaffected, since the decision criterion remains at the intersection of the two distributions.

(ii) As response optimization. A decline in signal detectability under the present circumstances (equiprobability of signals and nonsignals and a symmetrical payoff matrix) is an instance where response regulation makes the same prediction as response optimization: the decision criterion will always be placed at that position which corresponds to the optimal value of β of 1.0. Response optimization, which does not specify any feedback mechanism, predicts this as the result of a decision rule to adopt the optimal value of the likelihood ratio criterion. As mentioned in the previous section, the observer cannot transform any sensory observation into an accurate value of likelihood ratio without information to establish from which distribution signals are presently being sampled. This would require some form of either external or internal feedback, and in the situation under study no external information is provided to the observer. As a result, some sort of internal feedback must be employed to indicate when a decline in detectability has taken place. Whatever this 'cue', the outcome of response optimization will be to adopt an optimal likelihood ratio criterion, as is also predicted on the basis of response regulation.

(c) Response stabilization

If continuous response regulation takes place in this situation, there will be no reason for the intervention of the response stabilization process. That is, if the decision criterion is maintained at its "equal-odds" position, there will be no change in the proportion of observations falling on either side of it, and the ratio of signal to nonsignal responses will remain constant. However, if there is an initial tendency to maintain a fixed position of the decision criterion, then the response stabilization mechanism will demand an adjustment in its position at some point after the decline in signal detectability begins. This will follow from the movement to the left of

the signal distribution which will result in a decline in both the local, l , and cumulatively-based, c , probability of a signal response. Since the rate of decline will be greater for local probability, $(l-c)$ should take on a negative value. The continued decline in detectability should serve to further increase the negative difference between l and c , so that the expected movement in the criterion is the same as for a ramp decline in signal probability (see Figure 17), i.e. a progressive decline in the criterion. The criterion, however, will not be shifted so far as to be placed at its original "equal-odds" position, since revision in the observer's estimate of cumulatively-based signal probability will allow a smaller proportion of signal responses to result in a matching of local and cumulatively-based signal response probability. This adjustment is illustrated in Figure 42.

The expected trends in performance associated with response stabilization are somewhat more complicated than the previous sets of predictions since three eventualities can arise while response stabilization takes place: the downward rate of adjustment of the decision criterion during any period can be greater than, the same as, or less than the rate at which the signal distribution moves downward. However, an average rate of less than that of the signal distribution seems the most likely finding, since a decision criterion that moves with the same or greater average speed will ultimately be shifted to some point to the left, rather than right, of its original position at the intersection of the signal and nonsignal distributions.

Given the assumption that during response stabilization the decision criterion will move to the left along the sensory continuum more slowly than the signal distribution, the predicted patterns of performance are shown in Figure 43. During the initial period in which a fixed sensory cutoff is maintained, there will be a decline in the proportion of correct detections

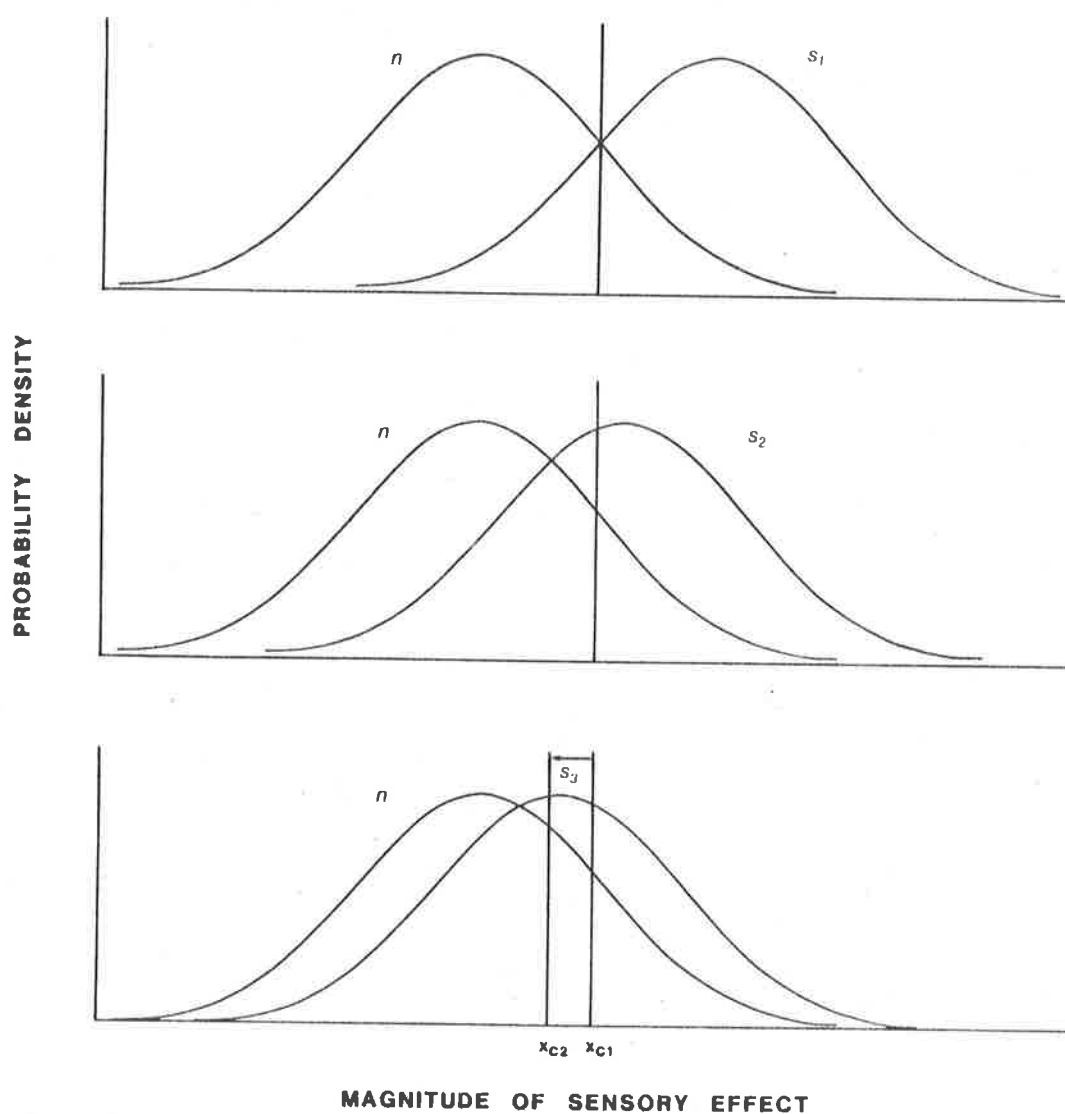


FIGURE 42. The adjustment in the observer's decision criterion expected by the response stabilization model to follow from an initial tendency to maintain a fixed sensory cutoff. At s_2 the observer maintains the cutoff, x_{c1} , that was adopted for s_1 . The decline in both l and c then lead to an adjustment toward a cutoff position, x_{c2} , to the right of the intersection of s_3 and n .

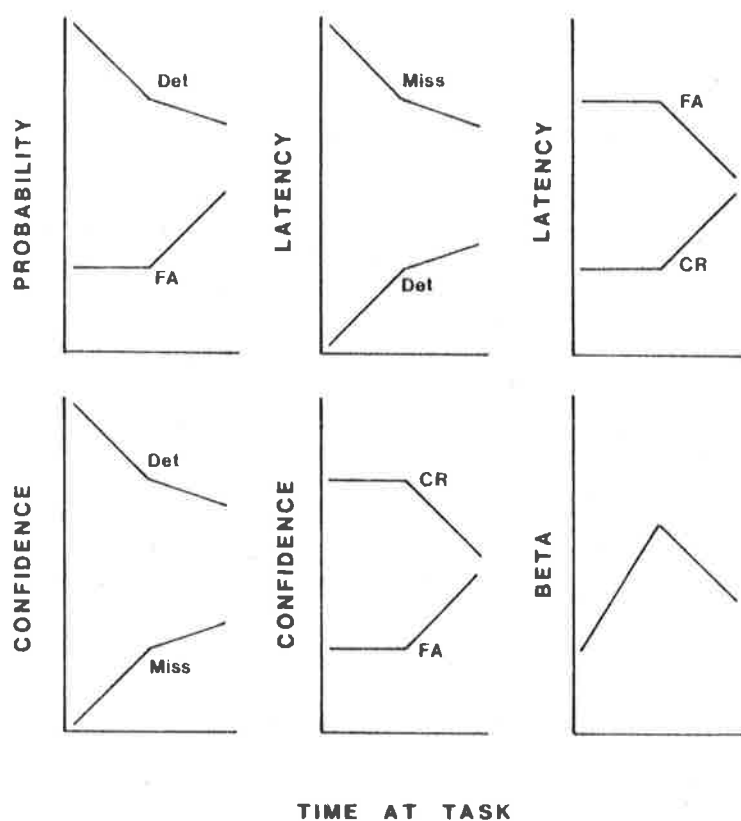


FIGURE 43. Patterns of change in performance that are expected to follow from an initial tendency to maintain a fixed cutoff followed by a tendency toward response stabilization, as shown in figure 42. The symbols are the same as those used in figures 39 and 41.

and an increase in the proportion of missed signals, as the signal distribution moves past the cutoff position. This will lead to decrease in the speed and confidence of correct detections, and an increase in the speed and confidence of missed signals. Since the decision criterion remains fixed with respect to the nonsignal distribution, there will be no change in the proportion of correct rejections and false alarms, and therefore no change in the latencies and confidence of these reports.

As the decision criterion is shifted to the left, following the intervention of the response stabilization process, but at a slower rate than the leftward movement of the signal distribution, there will be a reduction in

the rate at which detection probability declines and missed signal probability increases. This should lessen the previous tendencies for the speed and confidence of missed signals to increase. The movement of the criterion to the left with respect to the nonsignal distribution will now decrease the proportion of correct rejections, and increase the proportion of false alarms. This will decrease the speed and confidence of correct rejections, and increase the speed and confidence of false alarms.

The d' measure of sensitivity will show a progressive decline throughout the decrease in signal detectability, while the trends observed in β , as with the other response measures, will depend upon whether the decision criterion is fixed or is being adjusted. During that period in which a fixed position is maintained, the ordinate value of the criterion will increase only with respect to the signal distribution, so that the value of β will increase. During the period of response stabilization, the decision criterion will be shifted toward the intersection of the signal and nonsignal distributions from some position further to the right of the intersection, and the ordinate of the criterion on the nonsignal distribution will increase relative to its value on the signal distribution. The value of β will thus decline, but not far enough to reach its original "equal-odds" value.

The following experiment was designed to see which of the above sets of predictions would best describe performance when there is a progressive decline in the level of detectability at which signals are presented. There are clearly a number of differences in the predictions of the three models, and the main ones can now be summarized. All the models predict a decline in the probability of a correct detection, and an increase in that of a missed signal. This should be accompanied by a decline in the speed and confidence of correct detections, and an increase in the speed and confidence of missed signals. However, if response stabilization occurs, these trends should be

less pronounced after that point at which criterion adjustment is first initiated. Only the fixed sensory cutoff model predicts no change in any performance measure associated with false alarms and correct rejections. The response regulation model predicts that false alarm probability will steadily increase, while that of a correct rejection will decrease. As a result, there should be a continual increase in the speed and confidence of false alarms and a decrease in the speed and confidence of correct rejections. The response stabilization model also predicts these trends, but only after a period during which the probability, latency and confidence of these reports remain stable. Lastly, while all the models predict a steady decline in the signal detection measure of sensitivity, d' , they make different predictions about change in the measure of bias, β . A fixed sensory cutoff will produce a steady increase in this measure, whereas a response regulation process will result in β being maintained at its initial value, provided that an initial "equal-odds" value is adopted. The response regulation model predicts that if an increase in β is observed, it will be followed by a tendency for β to decline.

2. Method

(a) Task and stimuli

The detection task of Experiment 4 was again employed. On each trial the observer had to decide whether a 106 by 177 mm display consisting of horizontal lines which varied in length was a "signal" or "nonsignal". The signal was a display in which the lines in its central 84 by 106 matrix had a longer mean length than the lines in its perimeter (see Figure 30). The nonsignal display contained lines which had a constant mean length of 10.0 mm over the whole display, and a range of 7.1 to 12.9 mm. The lengths of the lines in the perimeter of the signal display had the same mean and range, but those in the central matrix were varied in order to produce seven

TABLE 15. Line lengths in the central matrix of signal and nonsignal displays. The lines in the perimeter of the displays always had a mean length of 10 mm. and range of 7.1 to 12.9 mm.

<u>Display</u>	<u>Detectability level</u>	<u>Line length (mm)</u>	
		Mean	Range
Signal	7	15.0	11.4 - 18.6
	6	14.3	11.4 - 17.1
	5	13.6	10.0 - 17.1
	4	12.9	10.0 - 15.7
	3	12.1	8.6 - 15.7
	2	11.4	8.6 - 14.3
	1	10.7	7.1 - 14.3
Nonsignal	All	10.0	7.1 - 12.9

distinct levels of detectability. The means and ranges of the line lengths in the central matrix are given in Table 15. Detectability level is numbered 1 to 7 in order of increasing detectability, i.e. the difference between the mean length of the lines in the centre and that of the lines in the perimeter.

The displays were presented on the screen of a television monitor, and viewed at a distance of approximately 4.4 metres. The observer was required to report the occurrence of both signals and nonsignals. Each type of report was followed by a rating of confidence along a 5-category scale to indicate degree of certainty in the accuracy of the primary decision response.

(b) Apparatus

Stimuli were generated by means of a graphics program contained in a D.E.C. PDP8/F computer. A second disk-based D.E.C. PDP8/E computer system determined whether each stimulus would be a signal or nonsignal and the length of time that the stimulus would be visible (2 seconds unless terminated by a decision response). Following stimulus termination there

was a further 2 second interval, which allowed a confidence response to be recorded, and which could also be terminated by a response. The second computer classified both responses, measured the latency of the primary decision response, and recorded the value of the confidence rating. Summaries of these statistics were printed after each block of training trials and at the conclusion of the main session. An interval of approximately 1.5 seconds followed the confidence interval before the next display appeared. This was the processing time required by the PDP8/E computer to determine the dimensions of the lines in the next display. The main session of 1000 trials could thus last for a maximum of about 92 minutes.

Observers depressed the left or right of two perspex buttons, lying on either side of a wedge-shaped resting position, to indicate whether a non-signal or signal, respectively, was judged to be present. Following this, confidence was expressed by depressing one of a line of ten perspex buttons, divided in the middle by a yellow line so that five were in front of each primary response button. The confidence scale for both responses ran from "least confidence" at the innermost button lying next to the yellow line, to "most confidence" at the outermost button.

(c) Design

The main session of 1000 trials was preceded by four blocks of 50 training trials each. Signals were presented randomly but with a constant probability of 0.5 throughout training and the main session. The levels of detectability at which signals were presented during the four blocks of training trials were 7, 4, 1 and 4, respectively, as shown in Figure 44. Thus training gave experience at the two extreme levels of detectability as well as the median level. The main session was divided into twenty blocks of 50 trials each. Detectability was varied during the session by gradually

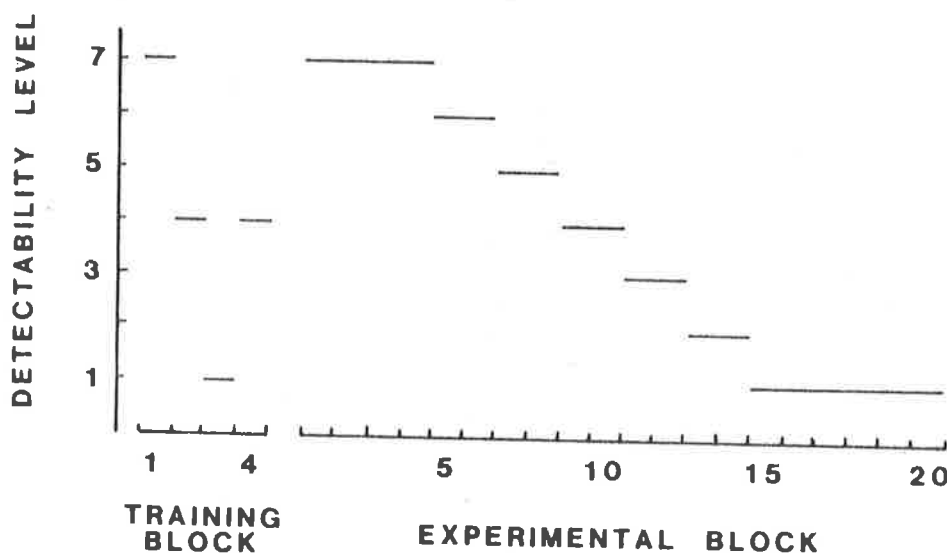


FIGURE 44. The level of detectability at which signals were presented during the four blocks of training trails and the twenty succession blocks of experimental trials in Experiment 5.

reducing the level of detectability from an initial level of 7 to a final level of 1 in the manner shown in Figure 44. Hence the level of detectability was 7 during trial blocks 1 to 4, 6 during blocks 5 and 6, 5 during blocks 7 and 8, 4 during blocks 9 and 10, 3 during blocks 11 and 12, 2 during blocks 13 and 14, and 1 during blocks 15 to 20. There were no breaks between trial blocks, and because of the small difference in stimulus magnitude between successive levels of detectability, the observer effectively monitored a continuous sequence of trials in which no abrupt change in detectability could be perceived.

(d) Procedure

Observers were tested individually and in complete isolation. Watches were removed prior to the first block of training trials. Initial instructions explained the nature of the signal detection task and an example of each type of display was shown to the observer. The use of the response buttons

and confidence scale was then explained, as was the need to make both responses within 4 seconds of stimulus onset and to always use the index finger of the dominant hand. Observers were told that task difficulty would vary from a level at which signals were very easy to detect to one at which signals were quite difficult, but they were not told the exact manner in which detectability would vary during the main session.

Half the observers (group I) were told that the probability of a signal would always be 0.5, both during practice and the main session to follow. This was mentioned again prior to the main session. The other half (group N1) were told that signals and nonsignals would appear with the same frequency during training trials, but no mention was made of their relative frequency in the main session. When questioned after the experiment, only one of the twelve observers in this latter group reported that she thought by these instructions that there "might" also be equiprobability during the main session. The remaining observers all made statements to the effect that they "really didn't think about it". These different instructions regarding signal probability were given in order to determine whether the explicit information that signals would always be as probable as nonsignals would affect any tendency which might be shown for signal response probability to change as signal detectability declined.

Prior to each block of training trials the observers were told that the signals to be presented would be "easy", "difficult", or of "medium difficulty", depending on the actual level of detectability to be employed. Following each block of training trials a print-out of performance data was checked to determine whether the observer was responding correctly. She was then told that her performance was satisfactory, and asked to continue to respond in the same manner.

(e) Observers

Twenty-four females enrolled in first-year Psychology at the University of Adelaide volunteered to participate in the experiment as part of their course requirements. They were randomly assigned to the two instruction groups. Ages ranged from 17 to 19 years with a median age of 17 years. All observers had normal or corrected vision.

3. Results

(a) Response probability

For the purpose of statistical analysis, the data of each observer were classified according to the two main factors of level of detectability and instruction condition. The results of analyses of variance on response probabilities are given in Table 16. No significant effect due to instructions emerged in any probability measure, although group differences were definitely indicated in the data. This was probably due to a high variation in performance among the individuals within each group, so that high within-groups error terms resulted. This was true not only for response probabilities but also for the other response measures to be reported. Such a high degree of variation among individuals in their performance in a detection task is not uncommon, and a separate analysis for instruction effects on overall performance will be reported in a later section.

(i) Conditional signal response probabilities. The significant main effect of level of detectability on the conditional probabilities of a correct detection and false alarm, reported in Table 16, is quite evident in Figure 45. As expected, the probability of correctly reporting a signal declined as its detectability decreased. This was accompanied by an increase in the probability of incorrectly reporting a signal, but mainly over the four lowest

TABLE 16. Results of analyses of variance on response probabilities over all levels of detectability.

	<u>Source</u>	<u>d.f</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	Correct signal response $P(S/s)$				
	Instructions (I)	1	.595	2.280	
	S within I	22	.261		
	Detectability (D)	6	2.463	97.232	.00001
	I x D	6	.009	.360	
	D x S within I	132	.025		
(b)	Incorrect signal response $P(S/n)$				
	Instructions (I)	1	.040	.119	
	S within I	22	.342		
	Detectability (D)	6	.551	34.162	.00001
	I x D	6	.009	.610	
	D x S within I	132	.016		
(c)	Signal response $P(S)$				
	Instructions (I)	1	.081	.632	
	S within I	22	.128		
	Detectability (D)	6	.192	17.163	.00001
	I x D	6	.003	.319	
	D x S within I	132	.011		
(d)	Correct response $P(S/s \text{ and } N/n)$				
	Instructions (I)	1	.236	1.364	
	S within I	22	.173		
	Detectability (D)	6	1.314	138.247	.00001
	I x D	6	.005	.621	
	D x S within I	132	.009		

levels of detectability. The stability of false alarm probability which is indicated over the three highest levels of detectability provides evidence of a fixed position of the decision criterion during the initial decline in signal detectability. Since such stability was also indicated in other response measures during this period, separate analyses of variance were carried out using data only from the three highest levels of detectability to determine which performance measures may have failed to exhibit any mean differences

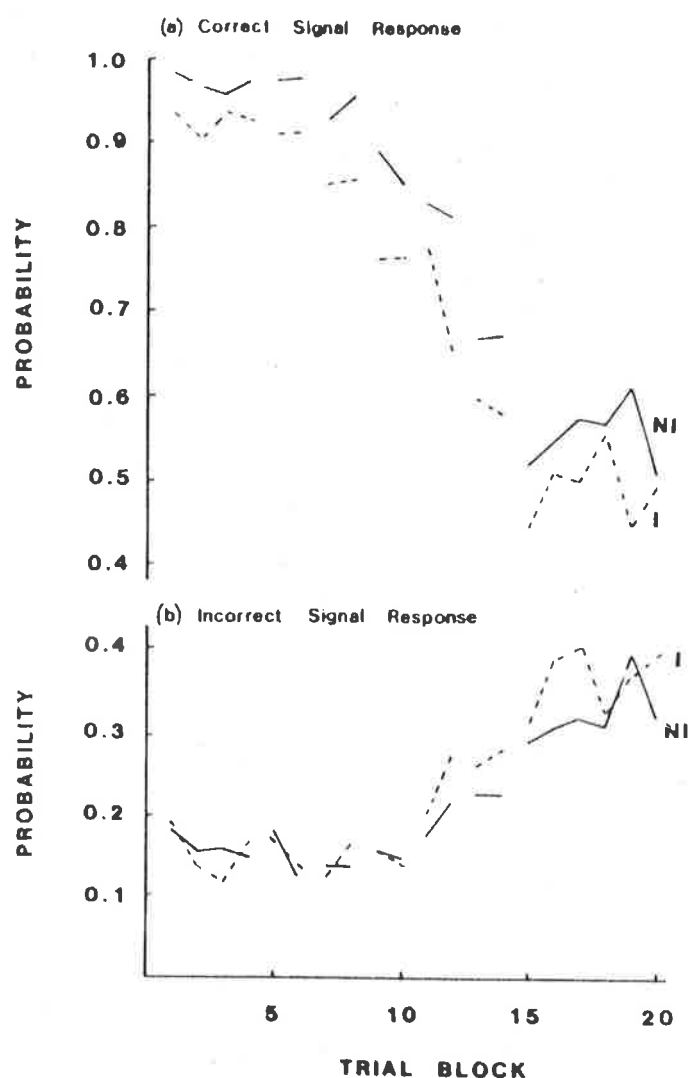


FIGURE 45. Changes in the mean conditional probability of (a) correct and (b) incorrect signal responses over the twenty trial blocks for each group of observers. Data points are joined across those trial blocks sharing a common level of detectability.

across these levels. The results of these analyses on response probabilities are given in Table 17, and confirm that detection probability declined while there was no change in false alarm probability across the three highest levels of detectability. Thus it appears that the cutoff position of the decision criterion along the nonsignal distribution initially remained fixed as the signal

TABLE 17. Results of analyses of variance on response probabilities over detectability levels 5 to 7.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	Correct signal response $P(S/s)$				
	Instructions (I)	1	.174	2.112	
	S within I	22	.082		
	Detectability (D)	2	.039	5.207	.01
	I x D	2	.008	1.077	
	D x S within I	44	.007		
(b)	Incorrect signal response $P(S/n)$				
	Instructions (I)	1	.0002	.001	
	S within I	22	.144		
	Detectability (D)	2	.004	.643	
	I x D	2	.001	.117	
	D x S within I	44	.006		
(c)	Signal response $P(S)$				
	Instructions (I)	1	.046	1.739	
	S within I	22	.026		
	Detectability (D)	2	.017	6.575	.005
	I x D	2	.001	.385	
	D x S within I	44	.002		
(d)	Correct response $P(S/s \text{ and } N/n)$				
	Instructions (I)	1	.040	.468	
	S within I	22	.086		
	Detectability (D)	2	.004	1.064	
	I x D	2	.003	.793	
	D x S within I	44	.004		

distribution moved to the left along the sensory continuum. The cutoff position was then adjusted toward the left over the four lowest levels of detectability, as shown by the increase in false alarm probability during this latter part of the task (Figure 45).

(ii) Unconditional signal response probability. The significant main effect of level of detectability on the unconditional probability of making a signal response was the decline shown in Figure 46. Contrary to original expectations, at the beginning of the task signal response probability was greater than a priori signal probability among both groups of observers. That it then declined to a value less than that of a priori signal probability indicates that criterion adjustments were not brought about in order to maintain signal

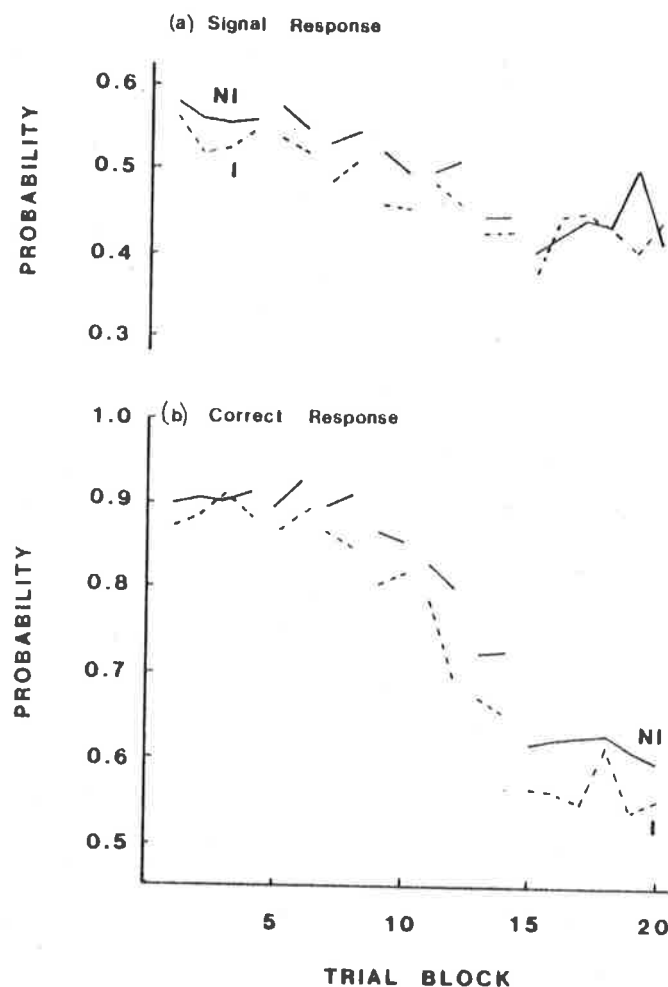


FIGURE 46. Changes in the mean unconditional probability of making (a) a signal and (b) a correct (signal or nonsignal) response over the twenty trial blocks for each group of observers. Data points are joined across those trial blocks sharing a common level of detectability.

response probability at the value of the observer's initial estimate of a priori signal probability, whether or not this estimate was accurate. However, Figure 46 does suggest that later in the task the criterion adjustments implied by false alarm probabilities may have at times served to counteract the decline in signal response probability. Specifically, it appears that signal response probability ceased to decline between detectability levels 4 and 3, then declined once more between levels 3 and 2, and, due to its increase over the lowest level of detectability, again ceased to decline between levels 2 and 1. These observations were confirmed by a significant comparison of mean signal response probability between detectability levels 3 and 2 ($F = 12.431$, d.f. = 1/22, $p < .005$) and nonsignificant comparisons between levels 4 and 3 ($F = .482$, d.f. = 1/22, $p > .4$), and 2 and 1 ($F = 1.211$, d.f. = 1/22, $p > .2$).

(iii) Correct response probability. As would be expected, Table 16 reports a strong main effect of level of detectability on the probability of making a correct response, signal and nonsignal responses combined. However, Figure 46 indicates that the decline in this measure occurred mainly over the four lowest levels of detectability, while a high level of accuracy was maintained over the three highest levels. Thus Table 17 reports that while the decline in the probability of a correct detection was great enough to produce a significant decline in the unconditional probability of a signal response over the three highest levels of detectability, the constant false alarm probability during this period resulted in a nonsignificant effect on the probability of a correct response.

(b) Response latencies

As was the case in the previous experiments, false alarms and missed signals proved to be too few to permit a thorough analysis of the latencies and confidence of these responses. It was only over the lowest level of detectability that the majority of observers were making both types of incorrect response, but neither the latency or confidence of either response exhibited any consistent trend or between-group difference. Therefore only the general trends associated with the latencies and confidence of correct responses will be presented.

(i) General trends. The results of analyses of variance performed across all detectability levels on the latencies of the two types of correct response are given in Table 18(a). The significant main effects of detectability can be seen in Figure 47 as increases in the latencies of both detections and correct rejections as detectability declined. Inspection of Figure 47 also indicates that the difference between the latencies of detections and correct rejections decreased, as detection latencies were initially lower but their increase with time was greater. This was confirmed by an analysis of variance of the latency difference score, the mean latency of correct rejections during each trial block subtracted from the mean latency of correct detections ($F = 18.904$, d.f. = 6/132, $p < .00001$ for the discriminability effect; $F = 1.394$, d.f. = 1/22, $p > .2$ for the instruction effect; and $F = .921$, d.f. = 6/132, $p > .4$ for their interaction). Such an effect could result if the signal distribution moved a greater distance toward the left past the decision criterion than the decision criterion later moved toward the left along the nonsignal distribution. The rate of increase in reaction times would then be greater for correct signal responses than correct nonsignal responses when testing across all levels of detectability.

TABLE 18. Results of analyses of variance on response latencies.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Over all levels of detectability:</u>				
(i)	<u>Correct signal response</u>				
	Instructions (I)	1	3.266	1.168	
	S within I	22	2.795		
	Detectability (D)	6	5.099	42.496	.00001
	I x D	6	.151	1.262	
	D x S within I	132	.119		
(ii)	<u>Correct nonsignal response</u>				
	Instructions (I)	1	.445	.116	
	S within I	22	3.829		
	Detectability (D)	6	1.091	14.130	.00001
	I x D	6	.056	.729	
	D x S within I	132	.077		
(b)	<u>Over detectability levels 5 to 7:</u>				
(i)	<u>Correct signal response</u>				
	Instructions (I)	1	2.994	2.447	
	S within I	22	1.223		
	Detectability (D)	2	.625	12.755	.00005
	I x D	2	.005	.119	
	D x S within I	44	.049		
(ii)	<u>Correct nonsignal response</u>				
	Instructions (I)	1	.291	.178	
	S within I	22	1.637		
	Detectability (D)	2	.009	.303	
	I x D	2	.012	.386	
	D x S within I	44	.031		

Table 18(b) reports that the detectability factor was significant only on correct rejections across the three highest levels of detectability. Figure 47 shows that the latencies of detections were increasing during this period while those of correct rejections remained constant. This agrees with the finding that the probability of a false alarm, and hence a correct rejection, did not change, while that of a detection decreased over the three highest

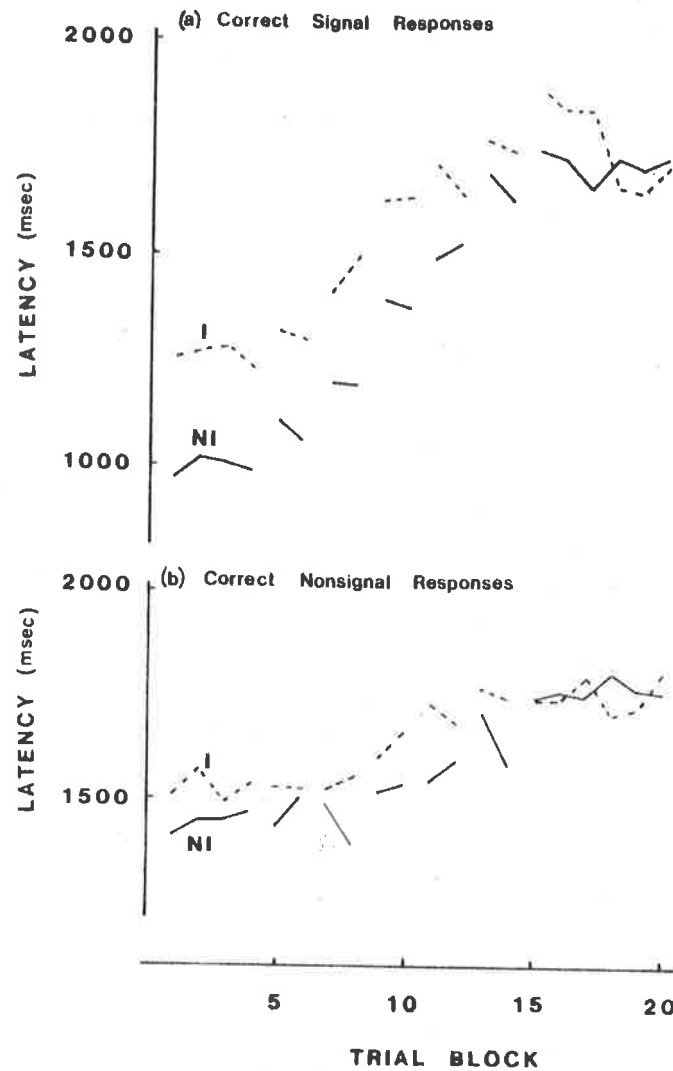


FIGURE 47. Changes in the mean latencies of (a) correct signal and (b) correct nonsignal responses over the twenty trial blocks for each group of observers. Data points are joined across those trial blocks sharing a common level of detectability.

levels of detectability. These results again indicate a fixed position of the decision criterion during the first phase of the experiment.

(ii) Differences among the four response categories. The mean latency for each of the four categories of response is given in Table 19. The means were calculated over the four lowest levels of detectability rather than the entire session as only during this phase of the experiment were there sufficient data for the inclusion of incorrect response latencies. Paired-sample t-tests confirmed that both types of correct response were made more quickly than their incorrect counterpart ($t = -9.24$, d.f. = 267, $p < .001$ for signals, and $t = -7.37$, d.f. = 272, $p < .001$ for nonsignals), detections were made more quickly than correct rejections ($t = -1.94$, d.f. = 285, $p < .05$), and there was no difference in latency between the two types of incorrect report ($t = -.61$, d.f. = 257, $p > .5$). The first finding is consistent with a distance-from-criterion interpretation, provided that the criterion is placed between the means of the signal and nonsignal distributions. The second finding implies that the criterion was usually placed closer to the mean of the nonsignal distribution over the four lowest levels of discriminability.

(c) Response confidence

(i) General trends. The results of analyses of variance performed across all seven levels of detectability on the confidence expressed in the two types of correct response are given in Table 20. The significant effects of detectability can be seen in Figure 48 as declines in both detection and correct rejection confidence as detectability declined. A confidence difference score was computed by subtracting mean correct rejection confidence during each trial block from mean detection confidence. There was a significant decline in this measure across the seven detectability levels and an interaction with instructions ($F = 24.095$, d.f. = 6/132, $p < .00001$, for the effect of detectability; $F = .004$, d.f. = 1/22, $p > .9$ for the instruction effect; $F = 2.891$, d.f. = 6/132, $p > .05$ for their interaction). Figure 48 shows that the decline

TABLE 19. Mean latencies and confidence for the four categories of response among all observers over detectability levels 1 to 4.

(a) <u>Response latency</u> (msec)			
<u>Correct signal</u> (S/s)	<u>Incorrect signal</u> (S/n)	<u>Correct nonsignal</u> (N/n)	<u>Incorrect nonsignal</u> (N/s)
1675	1863	1705	1839
(b) <u>Response confidence</u> (5-category scale)			
2.92	2.52	3.26	2.96

TABLE 20. Results of analyses of variance on response confidence.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Over all levels of detectability:</u>				
(i)	<u>Correct signal response</u>				
	Instructions (I)	1	6.906	.656	
	S within I	22	10.520		
	Detectability (D)	6	27.004	39.513	.00001
	I x D	6	.565	.828	
	D x S within I	132	.683		
(ii)	<u>Correct nonsignal response</u>				
	Instructions (I)	1	7.782	.777	
	S within I	22	10.012		
	Detectability (D)	6	3.107	12.609	.00001
	I x D	6	.565	2.293	.05
	D x S within I	132	.246		
(b)	<u>Over detectability levels 5 to 7:</u>				
(i)	<u>Correct signal response</u>				
	Instructions (I)	1	8.085	1.583	
	S within I	22	5.107		
	Detectability (D)	2	6.140	17.697	.00001
	I x D	2	.058	.168	
	D x S within I	44	.346		
(ii)	<u>Correct nonsignal response</u>				
	Instructions (I)	1	.800	.190	
	S within I	22	4.203		
	Detectability (D)	2	.001	.009	
	I x D	2	.977	5.447	
	D x S within I	44	.179		

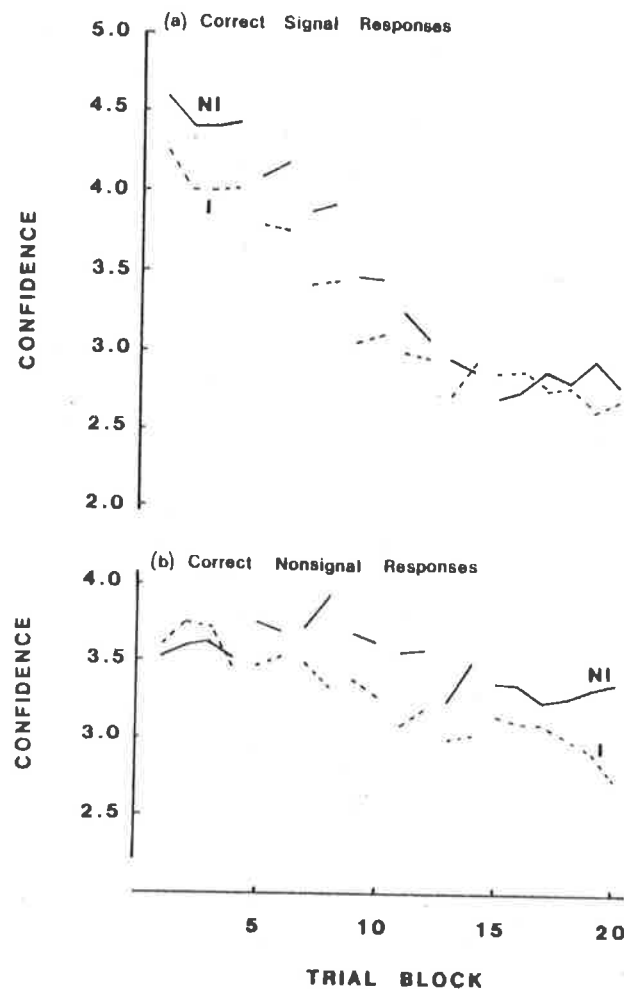


FIGURE 48. Changes in the mean confidence expressed in (a) correct signal and (b) correct nonsignal responses over the twenty trial blocks for each group of observers. Data points are joined across those trial blocks sharing a common level of detectability.

in the difference score was due to a greater rate of decline in detection confidence than correct rejection confidence. This parallels the finding that detection latency showed a greater increase with time than correct rejection latency, and might be explained in the same manner. The interaction can also be observed in Figure 48 as a greater rate of convergence of the two confidence measures among those observers who had not been

told that signals and nonsignals would always be equiprobable (group NI). This was probably the result of the rate of decline in correct rejection confidence being the least among these observers, which led to the significant interaction on correct rejection confidence reported in Table 20(a).

An analysis of confidence data solely from the three highest levels of detectability gave results quite similar to those obtained with response latencies. Table 20(b) confirms that detection confidence declined during this initial period of the experiment, in conjunction with the increase in detection latency, but there was no overall change in the confidence of correct rejections, as was also the case with their latencies. The interaction which was found on correct rejection confidence can be seen in Figure 48. Whereas the confidence of group I was decreasing over the three highest levels of detectability, that of group NI was increasing. In fact, the interaction which was found on correct rejection confidence across all seven levels of detectability can be attributed to these opposing initial trends. An analysis of variance on this measure performed across the four lowest levels of detectability yielded a significant effect of detectability ($F = 9.046$, $d.f. = 3/66$, $p < .0005$) but no effect of instructions ($F = 1.296$, $d.f. = 1/22$, $p > .2$) or their interaction. Thus it appears that at the beginning of the session those observers who were told that signals and nonsignals would always be equally likely (group I) exhibited a decline in their confidence in nonsignal responses as they became more frequent relative to signal responses, while the other observers (group NI) became more confident in the more likely response.

(ii) Differences among the four response categories. Mean confidence for each of the four categories of response is given in Table 19. As with response latencies these means were calculated across the four lowest levels of detectability. Paired sample t-tests confirmed that each type of correct

response was made more confidently than its incorrect counterpart ($t = 10.58$, $d.f. = 266$, $p < .001$ for signals, and $t = 8.68$, $d.f. = 271$, $p < .001$ for nonsignals), more confidence was expressed in correct rejections than in detections ($t = 7.57$, $d.f. = 285$, $p > .001$), and misses were made more confidently than false alarms ($t = 9.78$, $d.f. = 255$, $p < .001$). The first finding agrees with the result obtained with response latencies and is in accord with a distance-from-criterion interpretation of response latencies and confidence. However, the final two results, less confidence in signal reports, imply that the decision criterion was usually placed closer to the mean of the signal distribution over the four lowest levels of detectability, that is, at some position to the right of the intersection of the signal and nonsignal distributions. This is in direct contrast to the finding that detection latencies were shorter than correct rejection latencies during this latter phase of the experiment.

(d) Differences between instruction groups

(i) Within-task differences. In the preceding section there has been no report of any significant main effect of instructions on any response measure and in section 3(a) this was attributed to the variability in performance among the observers in each group. However, even a cursory glance at Figures 45 to 48 strongly suggest that between-group differences did exist. Table 21 gives the number of trial blocks in which each of the two instruction groups had the highest mean value of performance for each of the measures shown in Figures 45 to 48. It is possible, when within-group variability is ignored, to compute the binomial probability that either group by chance would have exhibited such an extreme number of instances in which it had the highest mean performance. These probabilities are also given for each performance measure in Table 21 and indicate that observers who were not

TABLE 21. The number of trial blocks in which each instruction group had the highest mean value of each of the performance measures shown in Figures 45 to 48. The binomial probability of the outcome is given for each performance measure.

<u>Performance Measure</u>	<u>Instruction Group</u>		<u>Binomial Probability</u> (2-tailed)
	<u>NI</u>	<u>I</u>	
(a) <u>Probability</u>			
Correct detection	20	0	.000002
False alarm	6	13	.17
Signal response	17	3	.003
Correct response	19	1	.00004
(b) <u>Latency</u>			
Correct detection	3	17	.003
Correct rejection	4	16	.01
(c) <u>Confidence</u>			
Correct detection	17	3	.003
Correct rejection	17	3	.003

told what relative probabilities to expect in the main session (group NI) tended to produce the higher mean probability of a correct detection, unconditional probability of a signal response and probability of a correct response. These observers also tended to make detections and correct rejections with lower mean latencies and higher mean confidence.

That observers in group NI tended to make more correct responses as the consequence of a higher detection rate indicates that the signal distribution was displaced further from the decision criterion among these observers. However, as there was no overall tendency for either group to have a higher false alarm probability, the position of the criterion relative to the nonsignal distribution was not influenced by instructions. Thus, instructions appear to have affected the distance between the means of the signal and nonsignal distributions, that is, the ability to discriminate between signal and nonsignal stimuli. This is supported by the tendency for observers

in group NI to make faster and more confident detection responses. Nevertheless, other factors may have been influenced, as is indicated by the tendency for observers in group NI to also make faster and more confident correct rejections even though observations leading to these reports should have been distributed at the same average distance from the criterion in both groups.

(ii) Pre-test differences. It seemed quite surprising that instructions about signal probability could somehow reduce the efficiency with which observers in group I detected signals. Therefore the hypothesis was tested that the differences observed between groups had existed prior to the experimental session. A check was made of the performance data obtained from observers during the last block of training trials prior to the main session, in which signals were presented at the median level of detectability (level 4). The differences in mean performance given in Table 22 confirmed that observers in group NI were better able to distinguish between signals and nonsignals. These observers had a higher mean detection probability ($t = 4.91$, d.f. = 22, $p < .0005$ (1-tailed)) combined with a lower mean false alarm probability ($t = -2.36$, d.f. = 22, $p < .05$ (1-tailed)) that corresponded to a higher group value of d' and lower group value of β . Although individual values of the signal detection measures could be obtained for only six observers in group NI and ten observers in group I, the difference between groups in d' was significant ($t = 2.06$, d.f. = 14, $p < .05$ (1-tailed)), while that in β failed to reach significance ($t = -1.75$, d.f. = 14, $p < .10$ (1-tailed)). The observers in group NI also made faster signal detections ($t = -2.32$, d.f. = 22, $p < .05$ (1-tailed)), and correct rejections ($t = -1.87$, d.f. = 22, $p < .05$ (1-tailed)) as appeared to be the case in the main session. However, unlike the differences indicated in the main session, the two groups did not differ in the mean

TABLE 22. Mean measures of performance for the two instruction groups during the final block of training trials.

<u>Performance Measure</u>	<u>Instruction Group</u>	
	<u>NI</u>	<u>I</u>
(a) <u>Probability</u>		
Correct detection	.95	.78
False alarm	.12	.21
(b) <u>Latency</u>		
Correct detection	1222	1602
Correct rejection	1344	1666
(c) <u>Confidence</u>		
Correct detection	3.88	3.68
Correct rejection	3.57	3.67
(d) <u>Signal detection</u>		
Sensitivity (d')	2.78	1.58
Bias (β)	.53	1.01

confidence of either detections ($t = .63$, d.f. = 22, $p > .25$ (1-tailed)) or correct rejections ($t = -.29$, d.f. = 22, $p > .25$ (1-tailed)). Hence the only effect of instructions on overall performance during the main session appears to have been on the level of confidence with which responses were made and, in view of the interaction effect found in correct rejection confidence, this probably resulted mainly from a tendency for correct rejection confidence to increase among observers in the NI group early in the task. Because of the random assignment of observers to the two instruction groups, rather than having the two groups equated in terms of some measure of efficiency of task performance, the initial differences between groups must be attributed to sampling error. This error, however, should not have affected any tendency which might have been shown for instructions to influence the way in which response ratios were revised as signal detectability declined.

(e) Signal detection measures and theoretical decision parameters

(i) General trends in d' and β . Although false alarms and missed signals were very uncommon over the highest levels of detectability, and so did not permit a signal detection analysis, the data of nine observers from group NI and all the observers of group I allowed the calculation of the signal detection measures over the four lowest levels of detectability. The results of analyses of variance on these data are given in Table 23. The sole significant effect, that of level of detectability on d' , can be seen in Figure 49. The constant decline in this measure confirms that the signal distribution was steadily moving toward the left along the sensory continuum during this phase of the experiment. On the other hand, there was no significant change in β , indicating that the decision criterion was being adjusted toward the left so that its relative position between the means of the signal and nonsignal distributions was little affected. As can be seen in Figure 49, the initial trend among both groups of observers was for β to increase toward, and then surpass, its optimal value of 1.0. This initial increase in the value of β is

TABLE 23. Results of analyses of variance on signal detection measures over detectability levels 1 to 4.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Sensitivity (d')</u>				
	Instructions (I)	1	3.201	1.314	
	S within I	19	1.886		
	Detectability (D)	3	11.904	58.116	.00001
	I x D	3	.053	.260	
	D x S within I	57	.204		
(b)	<u>Bias (β)</u>				
	Instructions (I)	1	.085	.036	
	S within I	19	2.363		
	Detectability (D)	3	1.344	1.657	
	I x D	3	.917	1.131	
	D x S within I	57	.810		

consistent with a signal distribution moving toward the left while the position of the decision criterion remains fixed, and is supported by the trends observed in the other response measures over the higher levels of detectability.

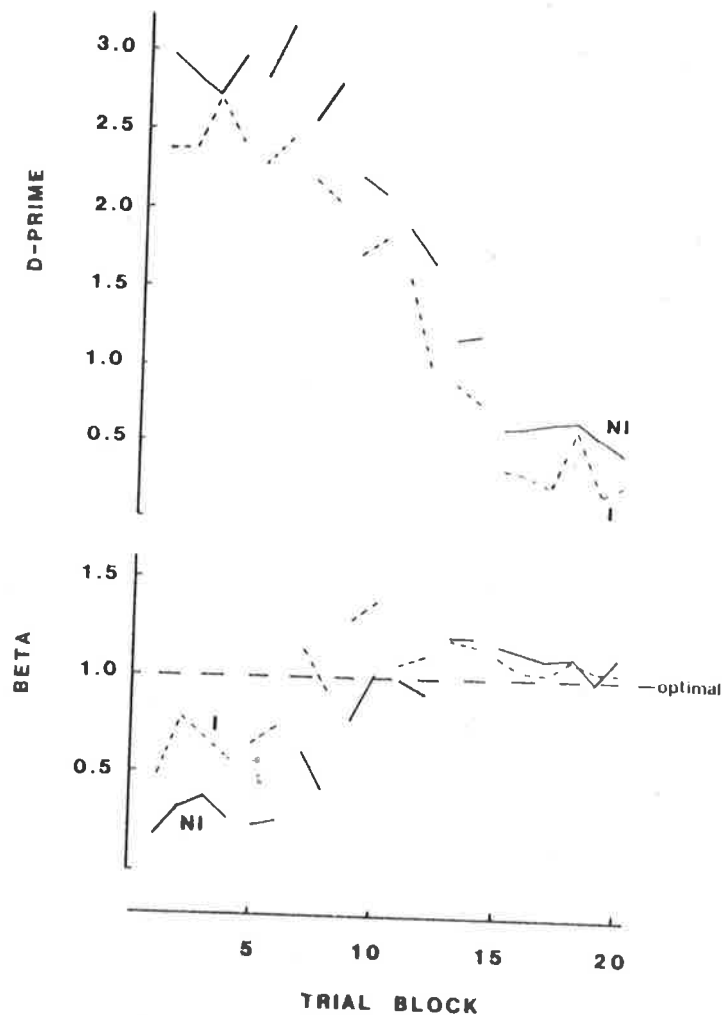


FIGURE 49. Changes in group values of d' and β over the twenty trial blocks for each group of observers. In the lower figure the values of β that were observed are compared with the optimal value (constant at 1.0) specified by signal detection theory. Data points are joined across those trial blocks sharing a common level of detectability.

(ii) Differences between groups. Figure 49 indicates that the differences shown between the instruction groups during practice in group values of d' and β continued on into the main session, despite the fact that there was no significant main effect of instructions on either measure. The NI group had the higher value of d' in each of the twenty trial blocks. The binomial probability that this result occurred due to chance is extremely low (.000002, 2-tailed). Also, the NI group had the lower value of β in each of the first twelve trial blocks, as was the case with the group values of β in practice. The binomial probability of this result is also low (.0005, 2-tailed). Over the remainder of the task the difference in group values of β were relatively small, although the NI group here had the higher value in seven of the eight trial blocks (binomial $p = .07$, 2-tailed). Contrary to what had been anticipated, neither group tended to adopt the optimal value of β of 1.0 initially, but instead adopted sub-optimal values. That group I had an initial β closer to the optimal value is as likely to have resulted from pre-task differences as from the instructions this group was given.

(iii) Relationship between β and $(1-c)$. As the same general trends were shown in both groups of observers, the data for the two groups were combined in order to compare changes in β with those in $(1-c)$, the difference between the local and the cumulatively-based probability of a signal response. Figure 50 shows group values of β , calculated in the usual manner from mean conditional signal response probabilities, and $(1-c)$, calculated from mean unconditional signal response probabilities, across the twenty trial blocks. Unlike the previous two experiments in which there was a stationary signal distribution, an increase or decrease in the value of β did not indicate an increase or decrease, respectively, in the decision criterion as defined by the amount of sensory evidence required to report the presence of a signal.

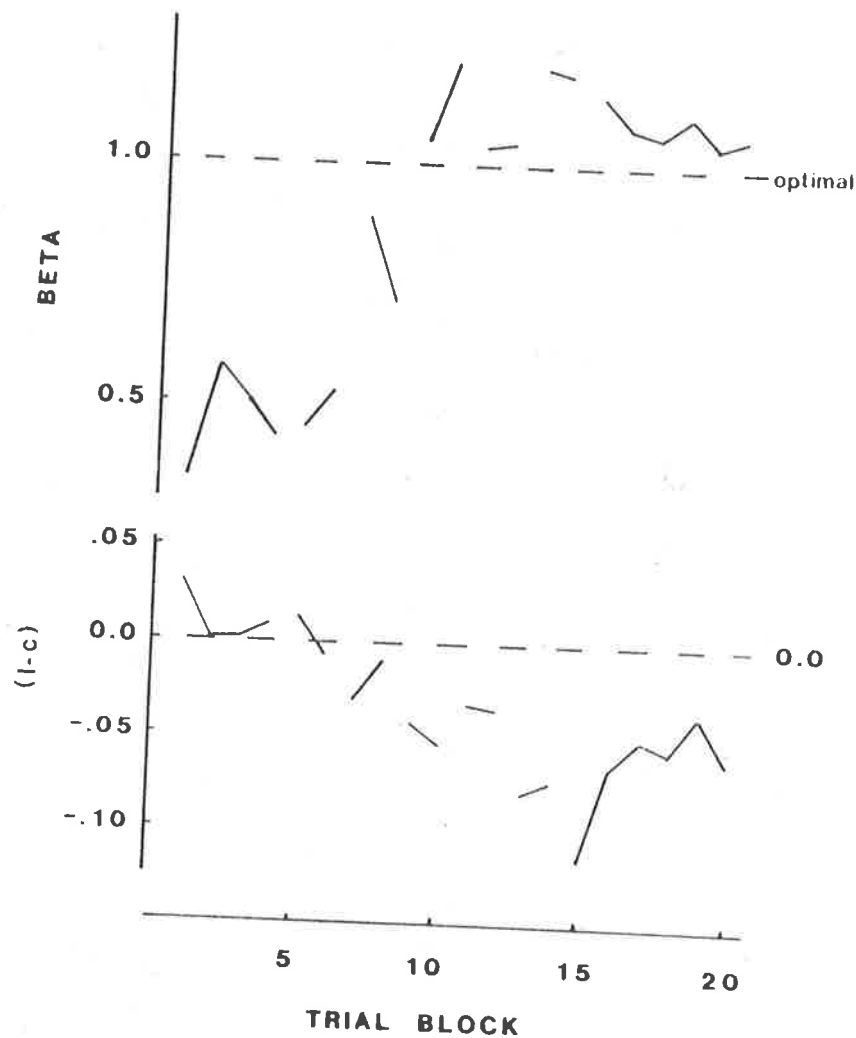


FIGURE 50 A comparison of changes over the twenty trial blocks in the value of β , calculated from the mean conditional signal response probabilities of all observers, and the mean value of $(1-c)$ for all observers. Data points are joined across those trial blocks sharing a common level of detectability.

Therefore, a measure of correlation between β and $(1-c)$ in the present experiment would be meaningless since change in β at any time may, or may not, have indicated a criterion adjustment.

However, Figure 50 shows there was a high correspondence between the adjustments made to the decision criterion, as are implied when change in β is considered within the context of the leftward movement of the signal distribution, and those adjustments specified by the response stabilization

model to directly follow the change in $(1-c)$. Over the two highest levels of detectability $(1-c)$ showed no consistent departure from zero, and there was no consistent trend in β . Then, $(1-c)$ declined from zero during the fifth and fourth levels of detectability, while the increase in β here indicated that the decision criterion was fixed as the signal distribution moved to the left. The decline in $(1-c)$, according to the response stabilization model, would have demanded a decline in the decision criterion, and the value of β did decline from the fourth to the third level of detectability, a change which could occur only through a leftward adjustment of the decision criterion. This resulted in a value of $(1-c)$ during the third detectability level that was higher than the fourth, so that an increase in the decision criterion was demanded or, alternatively, no change if the subsequent decline in the signal distribution produced a sufficient decline in signal response probability.

In either case, an increase in β would be expected, and in fact occurred from the third to the second level of detectability. As this produced a corresponding decline in the value of $(1-c)$, a final leftward adjustment of the decision criterion was demanded. As Figure 50 shows, a decline in β from its value in the second level of detectability did take place during the lowest level of detectability, indicating that the required criterion adjustment had occurred. This final adjustment served to once more increase the value of $(1-c)$ toward zero.

It should be noted that the initial and final values of β in the task indicated a criterion at first to the left and then to the right of the intersection of the signal and nonsignal distribution. This agrees with the prediction of the response stabilization model that, if adjustments of the decision criterion were required later in the task, the criterion would ultimately be placed at some position to the right of its original position

relative to the intersection of the signal and nonsignal distribution. This follows from the decline which took place in cumulatively-based signal response probability, which required a smaller recent proportion of signal responses to produce a zero value of $(1-c)$.

(iv) Indlin's (1976) index of β . Mean conditional signal response probabilities were used to compute values of the response regulation model's index of the decision criterion that was suggested by Indlin (1976) and is given in equation 6 above. Figure 51 compares the value of this index in each trial block with the value of β computed in the usual manner, both measures based on the combined data of both groups of observers. Two things are immediately obvious in Figure 51. Firstly, the values of the decision criterion specified by Indlin did not remain at some constant value throughout the task, thereby indicating that observers did not tend to maintain signal response probability at some fixed estimate of a priori signal probability. Secondly, the values of Indlin's index were always more radical than obtained values of β , being lower over the three highest levels of detectability, where the obtained values were less than 1.0, and higher over the four lowest levels of detectability, where the obtained values were greater than 1.0. Hence, if observers were adjusting their decision criteria in order to match local response probability with some estimate of a priori probability that varied over time, they showed consistent overestimation of the probability followed by consistent underestimation. The lower portion of Figure 51 shows that this tendency to overestimate and then underestimate signal probability occurred among both groups of observers, and so was not affected by information that signal and nonsignals would always be equiprobable.

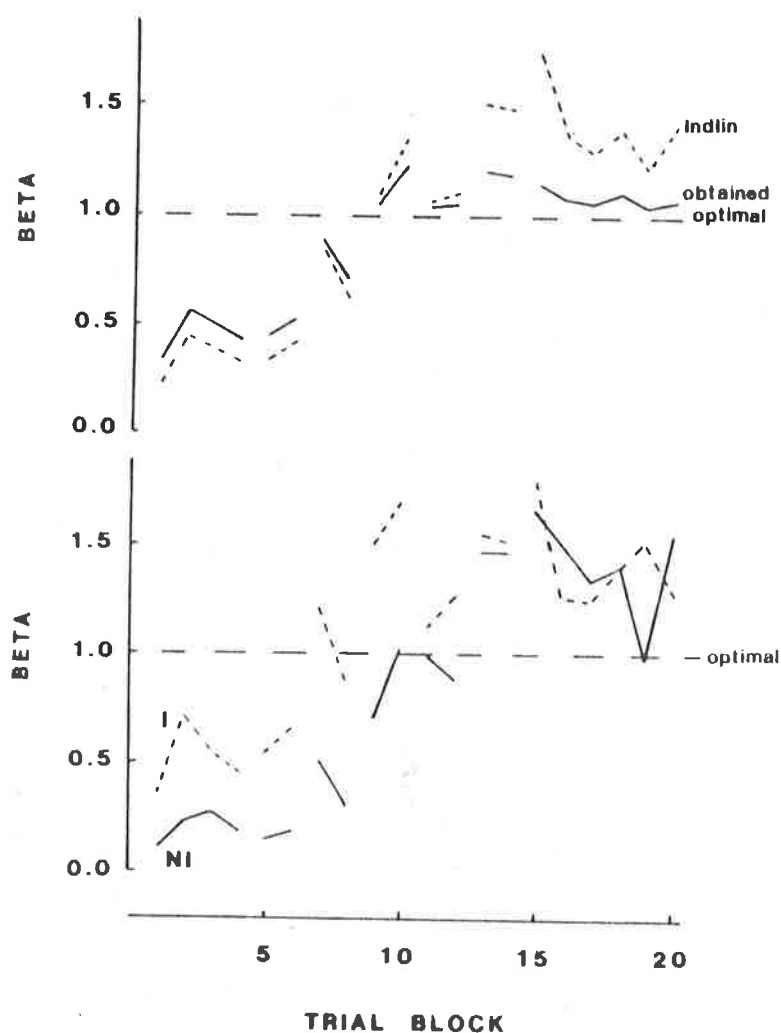


FIGURE 51 Values of the decision criterion index suggested by Indlin (1976) across the twenty trial blocks. In the upper figure Indlin's index is compared with the obtained group values of β and the measures are based on the data for all observers. In the lower figure values of Indlin's index are presented separately for each instruction group. Data points are joined across those trial blocks sharing a common level of detectability.

4. Discussion

(a) Relative support for the three models

(i) Fixed sensory cutoff followed by response stabilization. Of the three theoretical alternatives, the response stabilization model, combined with an initial tendency to maintain a fixed position of the decision criterion, clearly gained the greatest support. In line with the predictions of the fixed sensory cutoff model, there was a constant false alarm probability, combined with an increase in observed values of β , over the initial, higher levels of detectability. However, this tendency did not persist, as indicated by the subsequent increase in false alarm probability, and the relative stability in values of β over the later, lower levels of detectability. Both measures therefore reflected a decision criterion that was being adjusted downward along the sensory continuum. During this latter period of adjustment there was a high correspondence between the alterations in the decision criterion demanded by the response stabilization model, in terms of fluctuation in the value of $(1-c)$, and subsequent fluctuation in the value of β . These adjustments served to counteract the previous decline in signal response probability, but did not restore it to its initial level. They did, however, reduce the discrepancy between the local and cumulatively-based probability of a signal response.

(ii) Response regulation. The response regulation model predicted that values of β would remain invariant throughout the task. However, this prediction was based on the assumption that signals and nonsignals would be reported with the same frequency early in the task, so that an initial "equal-odds" β of 1.0 would be observed. That observers instead made an overly-high proportion of signal responses at the beginning of the task therefore indicates either that estimates of a priori signal probability were too high,

or that the proportion of signal responses exceeded estimates of signal probability. Either alternative would have led to specific consequences if response regulation had occurred. Firstly, if the estimate of a priori signal probability, as the reference signal, was too high, then signal response probability, as the feedback signal, should have been maintained at that high level throughout the task. However, signal response probability declined, showing that the reference signal had not been set at too high a value. Secondly, if the proportion of signal responses at the beginning of the task was higher than the observer's estimate of a priori probability, there should have been an immediate adjustment of the decision criterion in order to lower the proportion of signal reports. This would be especially true for those observers who were given information that signals and nonsignals would always be equiprobable. However, neither group of observers showed any tendency to alter the position of their decision criterion over the initial phase of the task. Hence, the decision criterion was not adjusted so as to maintain the proportion of signal responses at either the level of a priori signal probability, or the observer's initial estimate of that probability.

The only other explanation of the results on a response regulation model is that the estimates of a priori signal probability were initially too high, but were subsequently lowered, so that there was revision during the task in the value of the reference signal. However, since a priori signal probability itself did not vary, there is no justification within the framework of the model for variation in observers' estimates of this probability, particularly those observers who had been given information about signal probability. Nevertheless, the probability of a signal response fell to the same low level in both instruction groups.

In any case, the argument that the reference signal declined throughout the task serves only to convert the response regulation system into a response stabilization one. That is, the decline in the observer's estimate of a priori probability then becomes equivalent to a decline in the estimate of cumulatively-based probability in the response stabilization model, and the same adjustments would be predicted by both. Thus, on its own the response stabilization model can account for the present results, while the response regulation model must make additional assumptions which are inherent in the response stabilization model.

(b) Relationships among performance measures

The trends observed in the response latency and confidence measures of performance as detectability declined were in agreement with the distance-from-criterion hypothesis. As the probability of a response declined, implying that observations giving rise to that response were distributed more closely to the decision criterion, its latency increased and the confidence expressed in its accuracy decreased. This was found over the task as a whole for both correct detections and correct rejections. However, during the initial stages of the task, when the decision criterion remained at a fixed position along the nonsignal distribution, so that observations leading to nonsignal responses remained at an unchanged distance from the criterion, there was no change in the latency or confidence of correct rejections. During the same period, the signal distribution moved leftward past the criterion, so that observations leading to signal reports became distributed more closely to the criterion, and there was an increase in detection latency and decrease in detection confidence. Hence the changes observed in response latency and confidence were as valid indicators of criterion adjustment as were the changes in response proportions.

Differences among the four decision categories in response latency and confidence provide ambiguous support for a distance-from-criterion interpretation. In agreement with the interpretation, it was found that, for both signal and nonsignal decisions, correct response latency was less than incorrect response latency. Evidence conflicting with this hypothesis is provided by the finding that detection latency was less than correct rejection latency. This implies a criterion placed closer to the mean of the nonsignal distribution. At the same time, the confidence of detections was less than that of correct rejections, and the confidence of false alarms was less than that of missed signals. Both of these latter results imply a criterion placed closer to the mean of the signal distribution.

These differences were obtained over the final phase of the session, during which the overall value of β was always greater than the "equal-odds" value of 1.0, thereby reflecting a decision criterion placed to the right of the intersection of the two distributions, as implied by the data for response confidence. The finding that response latency did not exhibit the expected relationship with the distance between an observation and the decision criterion suggests that it might not be influenced solely by the distance from criterion. A possible additional influence is the "intensity component" of reaction time suggested by Gescheider, Wright, Weber, Kirchner and Milligan (1969), and later discussed by Pike (1973), which is indirectly proportional to stimulus intensity. Gescheider et al. found that when signal intensity increased, so that a greater proportion of the signal distribution was displaced to the right of a supposedly fixed decision criterion, the latency of false rejections increased at a lower rate than the rate at which latencies of correct detections decreased. They proposed that increased signal intensity decreased neural transmission time, thereby decreasing the intensity component

of the latency of responses based on observations arising from signal presentations, as well as increasing the proportion of the signal distribution lying to the right of the criterion. Hence, as signal intensity increased, the decrease in neural transmission time accentuated the decrease in detection latency and attenuated the increase in false rejection latency.

Such an intensity component of response latencies would continually decrease as we consider responses based on observations lying further to the right along the sensory continuum. As shown in Figure 52, there will be a progressive decline in response latency, as observations further to the right of the decision criterion are considered. This decline indicates a greater ease of decision-making as well as a decrease in the intensity component. Conversely, as responses further to the left are considered, response latency will first decrease, as decisions become easier, and then increase, as the intensity component increases. If such an intensity effect was present in the present study, then the latency of signal reports would be less than or equal to the latency of the nonsignal reports, even though the criterion was placed to the right of the intersection of the two sensory distributions.

On the other hand, if there is no such influence of sensory intensity on confidence, this measure will increase as a function of the distance that an observation lies from the criterion (see Figure 2). This means that the inverse relationship between response latency and confidence would not hold over the entire sensory continuum. While confidence in nonsignal responses will continually increase with the distance from the criterion, latency will at first decrease, and then increase, and so be directly related to confidence. Such a relationship would have to be verified experimentally.

The possibility that confidence is a more valid indicator of the ease of decision-making than latency might give some insight into the nature of

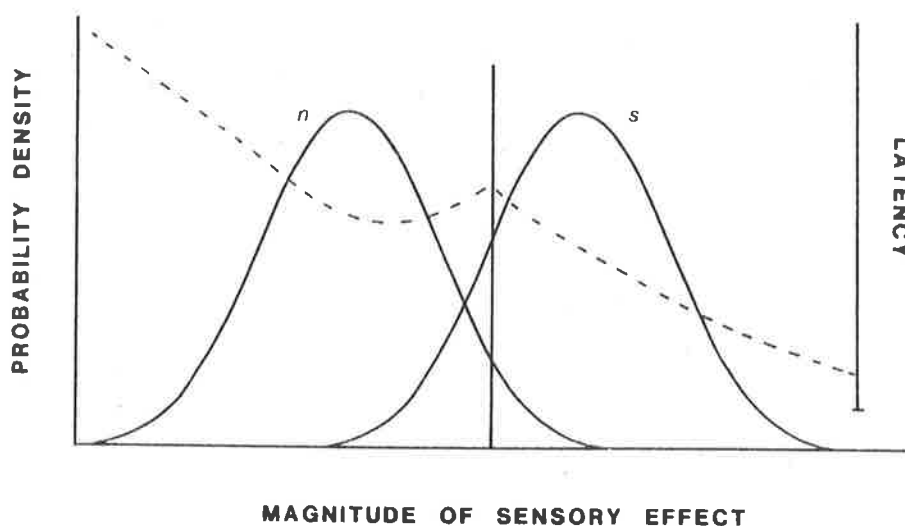


FIGURE 52 The relationship between response latency, given by the dashed line, and the distance of an observation from the decision criterion if there is an intensity component of response latency that is inversely proportional to sensory activity.

the only difference found between the two instruction groups which did not exist prior to the main experimental session itself. This was the higher level of confidence with which responses were made by observers who were not given probability instructions directly before the main session (group NI). During practice, group NI made a higher proportion of both correct signal and correct nonsignal responses, so that, theoretically, observations leading to both types of response were distributed further from the decision criterion. However, there was no significant difference between the two groups in the confidence measure. That confidence increased early in the task among observers in group NI, relative to group I, suggests therefore that observers were making confidence ratings more appropriate to the ease with which they were classifying events as either signals or nonsignals. It may be that observers tended to make intermediate ratings of confidence

during practice, but then made more extreme ratings when the highly detectable stimuli were encountered at the beginning of the experimental sequence. The greater tendency for confidence ratings to increase among observers in group NI would thus reflect the superior task performance of these observers.

(c) Relevance to previous studies in which detectability varied

Traditionally, when the effects of variation in signal detectability have been studied, performance has been compared between trial sequences differing in detectability, where the observer has been told of the particular level beforehand. Hence an appropriate decision criterion, whether in terms of likelihood ratio or amount of sensory activity, can be determined by the observer, and maintained throughout each trial sequence. On the other hand, Gescheider, Wright, Weber, Kirchner and Milligan (1969), Gescheider, Wright and Polak (1971), Gescheider, Wright, Weber and Barton (1971) and Curry, Nagel and Gai (1977) have varied signal detectability during each trial sequence of a task. The results of all these studies also strongly indicate that the observer maintains some fixed position of the decision criterion along the sensory continuum, consistent with the Neyman-Pearson objective of maintaining a constant false alarm rate. In the detection studies of Gescheider and his co-workers, signals of different intensities were presented randomly without the observer being told prior to any trial which intensity of signal could occur during that trial. Hence, it would have been impossible for the observer to adopt a different position of the decision criterion for each particular signal intensity. Although the stimuli used in the discrimination task of Curry et al. allowed the observer to determine which intensity level was in effect on each trial, the observer, in order to employ different criteria, would have had to adjust his decision criterion to an appropriate

position, as well as to evaluate the amount of evidence received in relation to that criterion position, during each stimulus presentation. Thus, the most convenient strategy again would have been to evaluate all stimuli in terms of the same sensory criterion.

The present experiment, in which signal detectability was gradually reduced rather than randomly varied, thus provided a situation in which the observer could utilize changes in the preceding response sequence as feedback in order to determine an appropriate position of the decision criterion prior to any trial. That false alarm rate increased during the latter part of the task then shows that there was no attempt to satisfy the Neyman-Pearson objective, which could have been accomplished by essentially ignoring changes in the previous response sequence and requiring a constant amount of sensory evidence in order to report a signal. Alternatively, if feedback was employed, as the results indicate, then, in order to satisfy the Neyman-Pearson objective, the observer would have required some record of response accuracy. The only record available to the observer in the present study was provided by the previous proportion of signal responses, and the criterion was adjusted in such a manner as to create stability between the recent and long-term proportions of those responses. However, such stability might have been the consequence of an attempt to maintain some desired reference level more related to the observer's conception of response accuracy than is given by the relative proportion of overt responses. If, as has been suggested, response confidence is a reliable index of the ease with which the observer can classify sensory events, then it may be that the observer adjusts his decision criterion in order to maintain confidence at some desired level, and thereby control the relative difficulty of his decisions. A model based on this possibility will be described in the next chapter.

CHAPTER 5.

A. A CONFIDENCE-BASED FEEDBACK MODEL

1. General Properties of the Model

The previous feedback models which have been considered assume that the observer keeps two estimates of signal probability, and that, whenever there is an excessive amount of discrepancy between the two estimates, there will be some adjustment to a decision criterion of the type conceptualized within signal detection theory. However, a somewhat different model can be developed, still within the framework of signal detection, if one follows the suggestion of Vickers, Leary and Barnes (1977) that the observer adjusts his decision criterion "so as to ensure that responses are made, on the average, with a certain target level of confidence" (p. 703).

There seems to be a number of advantages in specifying response confidence, rather than other response measures, as the reference signal in a feedback network. The previous experiment indicated that response confidence might be a valid index of the ease with which the observer classifies sensory events since it more closely reflected the distance an observation lies from the conceptual decision criterion than did response latency. Further, Vickers (1979, Chapter 6) has demonstrated the potential of response confidence as a source of information to the observer concerning the accuracy with which responses are made. Thus response confidence could provide the observer with more reliable information about the appropriateness of the decision criterion that he employs than is provided by response ratios, variation in which might in no way affect either the ease or the subjective accuracy of decisions. Lastly, by making criterion adjustments on the basis of changes in response confidence, the observer is not required to maintain a store of past events in order to calculate probability estimates, a feature of the

response stabilization model which has been criticized by Craig (1979).

Given that the degree of confidence in a response is directly related to the distance from the decision criterion of the sensory observation giving rise to that response, then the hypothesis that the observer attempts to make responses with a certain target level of confidence is analogous to the premise that the observer attempts to make observations at some constant average distance from the decision criterion. The average distance for each response is arguably that which is associated with the level of confidence established in that response during the initial training phase of a task. During training the observer is influenced to adopt a position of the decision criterion appropriate to the particular signal probability and payoff matrix employed, which could therefore lead to bias in the target levels of confidence in responses. For instance, if the observer adopts a high criterion when signal probability is low, the level of confidence associated with nonsignal responses will be greater than that associated with signal responses, since observations leading to nonsignal reports will fall further from the criterion. Hence, the target level of confidence for nonsignal reports will be greater than that for signal reports. On the other hand, if no biasing factors exist, and an equal-variance detection model is appropriate, then, by adopting a criterion at the "equal-odds" position between the means of the signal and nonsignal distributions, the observer will make signal and nonsignal reports at equivalent levels of confidence, and will attempt subsequently to maintain observations leading to signal and nonsignal reports at the same average distance from the criterion.

After initial learning of the task, any unexpected influence which results in an increase or decrease in the level of confidence with which responses are made will result in an adjustment of the decision criterion in

an attempt to restore confidence toward its previously established (target) levels. A necessary assumption of such a model, in order for it to be applied to the area of vigilance, is that there are target levels of confidence associated with covert responses, that are functionally equivalent to those associated with overt responses. Although the observer is not usually required to make nonsignal responses during a vigilance task, he nevertheless must make a signal or nonsignal decision on each trial. In terms of the model, the confidence with which a decision is reached will be the same as that with which a subsequent overt response is executed, and will have the same function in the regulation of ongoing performance. Thus the practice will be continued of using the term "response confidence", although it will be assumed to be that confidence with which a decision is reached, irrespective of whether it is followed by an overt response.

2. Determination of Mean Observation Distances

By use of the properties and tables of normal distribution given in Pearson and Hartley (1966) it is possible, given a set of observed or theoretical conditional response probabilities, to compute accurate average distances at which observations giving rise to each type of response lie from the decision criterion (the "mean observation distance" of each response). The distance from zero, ignoring sign, in standard deviation units to the mean of the greater portion of a normal distribution bisected by a criterion x_c is given by

$$g = \frac{Z(x_c)}{p(x_c)} \quad (8)$$

where $Z(x_c)$ is the value of the ordinate at x_c , $p(x_c)$ is the area in the greater portion of the distribution, and x_c is the absolute value, in standard

deviation units, of the position at which the criterion is placed along the distribution. Thus, the distance from the criterion of the mean of the greater portion of the distribution is given by

$$D(g) = \frac{Z(x_c)}{p(x_c)} + x_c \quad (9)$$

which is equivalent to the mean distance, in standard deviation units, that observations leading to a certain response lie from the decision criterion if $p(x_c)$ represents the conditional probability of that response. Conversely, for the alternative response, the mean distance that observations lie from the criterion in the lesser portion of the distribution is given by

$$D(l) = \frac{Z(x_c)}{1 - p(x_c)} - x_c \quad (10)$$

In terms of a signal detection model, the average distance from the decision criterion of all observations leading to a signal response is then given by

$$D(S) = \frac{p(S/s) p(s) D(S/s) + p(S/n) p(n) D(S/n)}{p(S/s) p(s) + p(S/n) p(n)} \quad (11)$$

where $p(s)$ and $p(n)$ are the a priori probabilities of signal and noise, respectively, $p(S/s)$ and $p(S/n)$ are the conditional probabilities of a detection and false alarm, respectively, $D(S/s)$ is the average distance from the decision criterion of the portion of the signal distribution associated with signal responses, and $D(S/n)$ is the average distance from the criterion of the portion of the nonsignal distribution associated with signal responses. The mean observation distance of nonsignal responses can be calculated in a similar fashion by using the complementary probabilities and their associated mean distances from criterion.

As can be seen in the preceding equation, for any value of the decision criterion, a change in a priori probabilities will affect the mean observation distance of the two responses. Mean distances are also related to conditional response probabilities, so that they will be affected by any change in the distance, d' , between the means of the two stimulus distributions. Thus, those manipulations which affect relative response proportions will also affect mean observation distances. Consequently, a signal detection model in which response confidence is used as feedback to determine when criterion adjustments are necessary can be applied to the same data as the response stabilization and regulation models, allowing them to be merged within a unified framework. This will now be illustrated by examining the predictions of the confidence-based model about the observer's response to variation in signal probability and signal detectability.

3. Response to Variation in Signal Probability

(a) Ramp change in signal probability

Figure 53 gives the mean observation distance of signal and nonsignal reports for values of signal probability between .5 and 0.0 when d' is 2.0 and the decision criterion is placed at the intersection of the signal and nonsignal distributions. The distances were calculated by use of the formulae presented in the preceding section. Due to the assumptions of symmetry and homogeneity of variance, the mean observation distances that would be associated with values of signal probability from .5 to 1.0 would be given by the curve for nonsignal responses over the complementary range of signal probability values that are given in Figure 53, and vice versa for the mean observation distances of nonsignal responses.

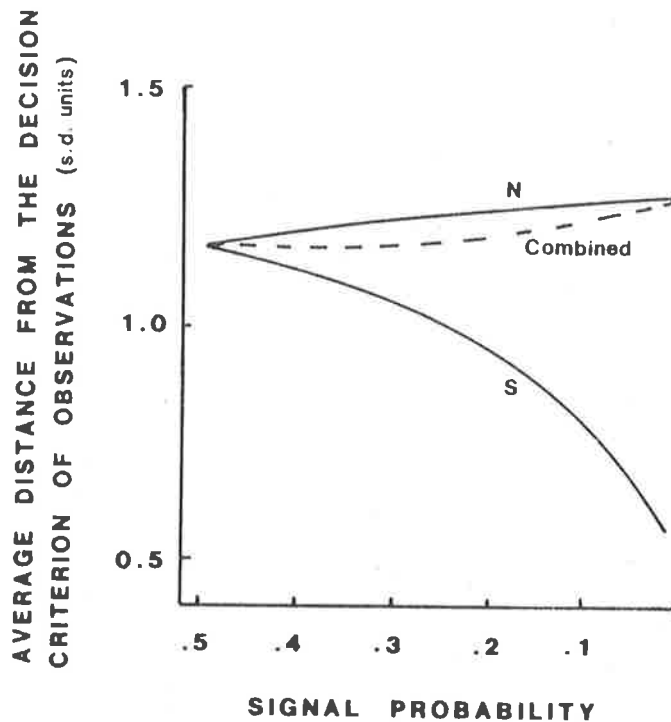


FIGURE 53 The average distance at which observations lie from the decision criterion as signal probability declines from .5 when d' is constant at 2.0 and the criterion remains fixed at the intersection of the s and n distributions. Observations leading to signal responses are denoted by S and those leading to noise responses by N . The combined average distance of all observations is given by the dashed line.

(i) Effects on signal and nonsignal response confidence. As is indicated in Figure 53, if signal probability declines from .5 and the decision criterion remains fixed, there will be a rather large decline in the mean observation distance of signal responses, and a less marked increase in that of nonsignal responses. Consequently, according to the model, confidence in signal responses will decline sharply, while confidence in nonsignal responses will show a tendency to increase. Such respective under-confidence and over-confidence relative to previous target levels can only be reduced by adjusting the decision criterion toward the left along the sensory continuum, thereby increasing the mean observation distance of signal responses and decreasing

that of nonsignal responses. A consequence of this adjustment will be an increase in the conditional probability of both types of signal report and a decrease in the signal detection measure, β , the results which were observed at each stage of the decline in signal probability in Experiments 1 to 4.

The converse effects would be predicted if signal probability is raised rather than lowered. The predicted response to a ramp change in signal probability is thus the same as that for the response stabilization model, in which it would be required in order to minimize the discrepancy between the local and cumulatively-based probability of a signal report. For similar reasons, the same response is predicted by the response regulation model. Again, however, there is strong disagreement with the response optimization model: if the observer were to adopt the "optimal" value of β at each value of signal probability given in Figure 53, the disparity between the curves for signal and nonsignal responses would be increased rather than decreased.

(ii) A possible "trade-off" between overall and relative response confidence. Also presented in Figure 53 is the weighted mean confidence of all responses at each value of signal probability. An increase in the general level of response confidence as signal probability declines results from the decrease in the weighting given to the combined mean by signal response confidence. Hence, by maintaining the decision criterion at the initial position the observer would in fact increase the general level of confidence with which responses were made. Even further increases would ensue if the decision criterion was always placed at the optimal location specified by signal detection theory. This is illustrated in Figure 54 in which the mean observation distances of signal and nonsignal reports, as well as their combined mean, are given for the data of Experiments 1 and 4.

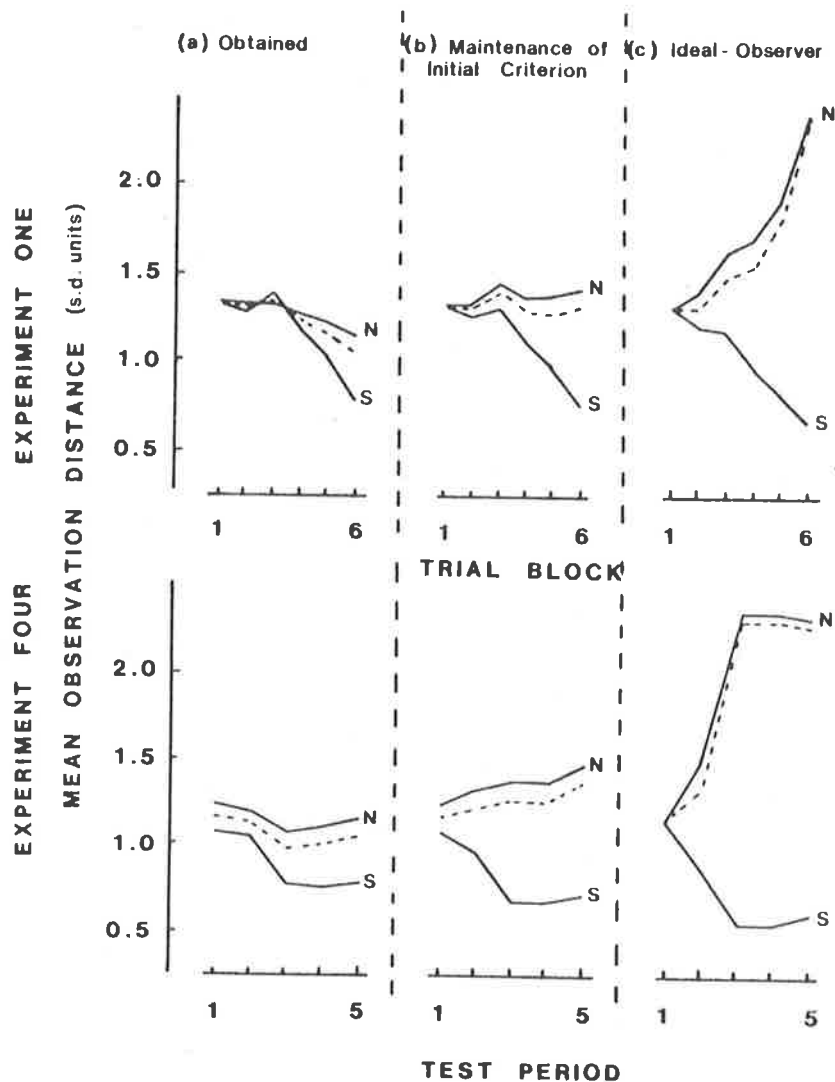


FIGURE 54 The mean observation distances of signal (S) and nonsignal (N) responses for each trial block of Experiment One and test period of Experiment Four. The obtained distances (a) are compared with those which would have resulted if (b) the initial criterion remained fixed or (c) the criterion was always placed at the optimal signal detection theory location. In each figure the combined mean distance of all observations is given by the dashed line.

These obtained observation distances are compared with those which would have resulted had the observer either maintained a fixed position of the decision criterion or placed the criterion at the optimal location specified by signal detection theory.

An inspection of Figure 54 suggests the possibility that there may be some sort of "trade-off" between the extent to which the observer will adjust the decision criterion, and so reduce the relative discrepancy between the confidence held in each type of response, and the amount to which the overall level of confidence is thereby reduced. In both experiments there was a decline in the general level of confidence with which responses were made. However, as the task progressed the imbalance between the amount of confidence with which signals and nonsignals were reported was much less than would have occurred if there had been no criterion adjustment, or if the criterion had always been placed at its optimal position. Since these latter two strategies would have sustained, or increased, the overall confidence of responses, it appears as if observers are prepared to tolerate some decline in the general level of confidence in their attempt to reduce the discrepancy between the relative confidence of responses.

(b) Step change in signal probability

The preceding section demonstrated that in response to a ramp change in signal probability the confidence-based model will take essentially the same action as the feedback models which employ response probability as a reference stimulus: an adjustment of the decision criterion in the opposite direction to the optimal location specified by signal detection theory. This will necessarily be the initial adjustment that is required when there is an abrupt step change in signal probability. In the literature, step changes in signal probability have been reported from training to experimental trials, as well as during the experimental sequence itself. These two situations will be treated separately.

(i) During the experimental sequence. In the vigilance task of Williges (1973), signal probability was increased from .1 to .5 midway through the experimental sequence. Although observers were informed beforehand of this change, there was no tendency for β to decrease toward its optimal value. Instead, a small increase in β was accompanied by a significant decline in detection probability. This strongly suggests that there had been an increase in the decision criterion, although the initial effect upon β may have been masked to some extent by the fact that data were averaged over blocks of 150 trials for analysis. Embrey (1975), on the other hand, did not inform observers in one condition that signal probability would be reduced from .5 to .2 during an industrial inspection task. In full agreement with the model, β declined from the value of 1.18, prior to the reduction in signal probability, to a value of .80 after the reduction.

(ii) From training to experimental trials. Williges (1969) found that observers whose training had led them to expect signals and nonsignals to be equiprobable, and who would therefore seek roughly equivalent levels of confidence in both types of decision, adopted a low value of β when tested at a low signal probability, and a high value of β when tested at a high signal probability. Observers undoubtedly started adopting an intermediate criterion, which was then rapidly adjusted (see Figure 3). The result was that confidence in signal responses was lowered toward its training level when observers were tested at the high signal probability, and increased toward its training level when they were tested at the low signal probability.

Parducci and Sandusky (1970) provided observers with no information regarding a priori probabilities prior to a task that required the identification of two tones which differed in amplitude and were presented at relative

probabilities of either .2, .5 or .8. If it can be assumed, as in Williges' study, that the lack of any biasing factor in training led observers to seek equivalent levels of confidence in their responses, then a high decision criterion would again be predicted at the high stimulus probability and a low criterion at the low probability. Although values of β were not reported by the authors, the measures of conditional response probability which were reported show that this was the case.

(iii) The possibility that target levels of confidence may be "overshot".

Unlike Williges (1969) who found β to be invariant, Colquhoun and Baddeley (1967) found that β decreased with time in a vigilance task when observers were trained at a low and tested at a high signal probability, while it increased when they were trained at a high and tested at a low signal probability. These trends (the 'recoveries' specified by the response stabilization model) are thus reversals in the initial adjustments that are demanded by the model. Although group differences over time in β were not reported, evidence that the initial adjustments had occurred appears in Figure 1 of Colquhoun and Baddeley (1967). When observers had been trained at a high signal probability, detection probability over the first block of 200 trials was highest among those observers who were tested at the low signal probability. Conversely, when observers had been trained at the low signal probability, detection probability over the first block of 200 trials was lowest among those observers who were tested at the high signal probability.

According to the model, a reversal in the direction of criterion adjustment would occur if the initial adjustment caused response confidence to be shifted beyond its target level, so that subsequent corrective adjustments would be required to restore it to that value. That is, there may be

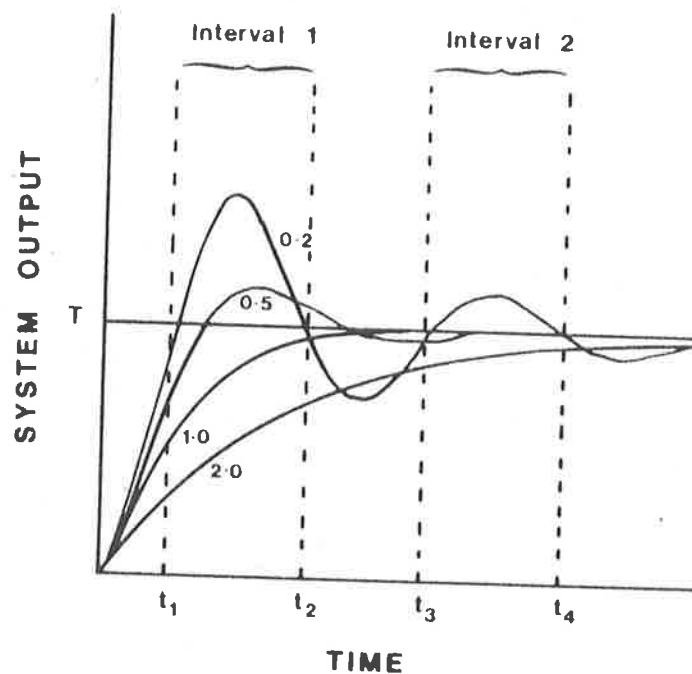


FIGURE 55 The rate at which the output of a second-order feedback control system approaches its target level (T) under various values of the damping ratio. The time interval from t_1 to t_2 is equal in duration to the subsequent interval from t_3 to t_4 .

a tendency for response confidence to "overshoot" its target level. Such a tendency is characteristic of second-order feedback systems which have low damping ratios so that in response to a step input to the system, system output tends to overshoot the desired input value rather than more gradually approach it (McFarland, 1971; Toates, 1975). Figure 55 shows the rate at which such a system responds to a step input (its target output) under various values of the damping ratio. When there is a "critical" damping ratio of 1.0, the system output just fails to overshoot its input value, when the ratio is greater than 1.0 (overdamped) there is a more gradual approach toward the input value, and when the ratio is less than 1.0 (underdamped) system output overshoots its input value, and then shows decaying oscillation toward that value.

The response of the confidence-based mechanism to the disparity between actual and target levels of confidence when there is a step change in signal probability might be analogous to the response to a step input of the above type of feedback system. If so, the manner in which response confidence approaches its target level could take on any of the patterns in Figure 55. The underlying adjustments in the decision criterion would necessarily exhibit the same pattern as the change in confidence but, in terms of the measure, β , be in the opposite direction to that of confidence. Hence, Figure 55 represents the possible patterns of change which might occur in the decision criterion in response to a step increase in signal probability (the figure rotated about its horizontal axis would then depict the responses to a step reduction).

All the patterns shown in Figure 55 indicate the same general trend, an increase in the criterion. However, in time-analyses of performance individual values of β are usually computed from data averaged over a fairly large sequence of trials, so that the initial overshoot of an underdamped system could lead to erroneous conclusions about the direction in which the criterion is adjusted. For instance, if, in Figure 55, the value of β computed in the interval from t_1 to t_2 is compared with the value in the interval from t_3 to t_4 , then β can be concluded to have either increased, remained constant, or decreased, depending on the value of the damping ratio. Colquhoun and Baddeley (1967) may have found their within-session trends in β because they tended to compare values of β obtained during an initial overshoot with values later in the task when the criterion was closer to its final steady-state value. The invariant values of β in Williges' (1969) study would then represent the effects of near-critical damping, in that β quickly approached its steady-state value without drastic overshoot.

One factor which might be related to the tendency to overshoot is the amount of information which is given to the observer about the changes in signal probability which are to occur in a task. Thus no information might lead to no damping effect, that is, the system's response is not restrained in any way since the observer is guided only by changes in confidence. Detailed and accurate information then might result in the greatest damping effect. As Colquhoun and Baddely (1967) gave observers no instructions about signal probability, the tendency to overshoot would have been greater than in Williges' (1969) study, where observers were given information, albeit inaccurate, about the signal probability to be expected in the task, and which could have been used to quickly establish whether levels of response confidence were appropriate. An even greater damping effect would have been present in Williges' (1973) study where observers were told of the change that would take place in signal probability, which could explain why there was little evidence of criterion adjustment following the change. Further, the combined effect of factors such as instructions, previous task experience, and ongoing knowledge of results could be great enough to totally suppress any response demanded by the system. Of course, with complete and accurate information about task characteristics, and external feedback about performance, there would be no need for the observer to utilize information based on confidence, as an implicit measure of response accuracy, in order to control ongoing performance.

4. Vigilance Decrement

The decline in detection rate often observed during the course of vigilance tasks has been shown to follow from an increase in the criterion for a signal response (refer to Chapter 1, section B. 4. (b)). Colquhoun and Baddeley (1964, 1967) have suggested that many of the reported cases of this vigilance decrement can be largely attributed to the practice of training observers at a signal probability that is higher than that at which they are tested. In the previous section, it was suggested that such increments in β might represent a slower upward adjustment in the decision criterion, following an initial adjustment that had caused signal response confidence to overshoot its target level. However, vigilance decrements can also be predicted when there has been prior training at the signal probability employed in testing. This will happen whenever observers adopt a more intermediate criterion at the start of a vigilance task than during training. That is, the value of β will be closer to the "equal-odds" ratio of 1.0. At the low signal probability, typically employed in vigilance tasks, this will result in the confidence of signal responses being increased, and that of nonsignal responses decreased, relative to the levels established during training. In order to restore response confidence to previous levels the decision criterion must therefore be increased.

The reason that observers who have been trained at the appropriate signal probability would adopt a more intermediate decision criterion might result from a phenomenon similar to an increased "expectancy" due to lack of previous task experience. That is, unless prior training is very rigorous, there may be a trial-and-error process at the beginning of a vigilance task by which the observer first adopts a decision criterion somewhere around the middle of the range of sensory activity that was experienced in training,

so that it is measured as a β closer to the value of 1.0. Then, if necessary, the cutoff value of the criterion is increased or decreased to that value which results in the approximate matching of response confidence with its target levels. Hence, when no change in signal probability occurred between training and testing, observers in Williges' (1969) experiment adopted a decision criterion corresponding to a β of about 1.5 at the beginning of a vigilance task whether they were tested at a probability of .17 or .83. The within-session trends that followed were for β to increase at the low signal probability, and decrease at the high. If the decision criterion adopted in training was appropriate to a priori probability, this would have restored response confidence toward its training levels.

Irrespective of the reason that a more intermediate criterion might be adopted early in a task, the model would predict vigilance decrements to be larger where this tendency is greatest. Indirect evidence that this is so is given by Teichner (1974), who analyzed the data from thirty-seven vigilance studies, and plotted detection probability as a function of time on watch as well as the initial level of detection probability in the task. Although no analysis was made of false alarm probability or signal detection measures, all of the studies under examination would have employed signal probabilities of less than .5. As a result, initial detection probability can be considered to have been closely and directly related to the degree to which observers tended to adopt an intermediate criterion at the beginning of a task rather than a more appropriate high value of the decision criterion. Consistent with the model it was found that the degree to which detection probability declined with time was greatest at the highest initial level of detection probability (.90 to .99) and there was no decline in detection probability at the lowest initial level (.30 to .39).

Lastly, it should be noted that Baddeley and Colquhoun (1969) found that β was higher at all times during a vigilance task than it was during training. However, performance measures were calculated over sequences of 200 trials, so it is not certain whether initial values of β would have been higher than during training if they had been calculated over a shorter series of trials. Increases were also indicated in d' between training trials and the experimental session among the five groups of observers in the study. Such increases in signal detectability may occur through a learning process that has not been completed by the time the experimental sequence begins. Under these circumstances, the model would again predict an increase in β . This prediction will be illustrated in the following section.

5. Response to Variation in Signal Detectability

Variation in signal detectability will alter the mean observation distance of signal and nonsignal responses in a manner that will again demand specific criterion adjustments according to the confidence-based feedback mechanism. Figure 56 (a) gives the mean observation distance of the two responses as the separation between the means of the two stimulus distributions is increased while the decision criterion remains at a fixed position one standard deviation unit above the mean of the nonsignal distribution. There is a rather sharp increase in the mean observation distance of signal responses as the mean of the signal distribution becomes displaced further from the criterion, accompanied by a shallow increase in the mean observation distance of nonsignal responses. The latter effect is due to the decline in the relative proportion of the signal distribution falling below the criterion, that is, observations lying close to the criterion leading to nonsignal responses.

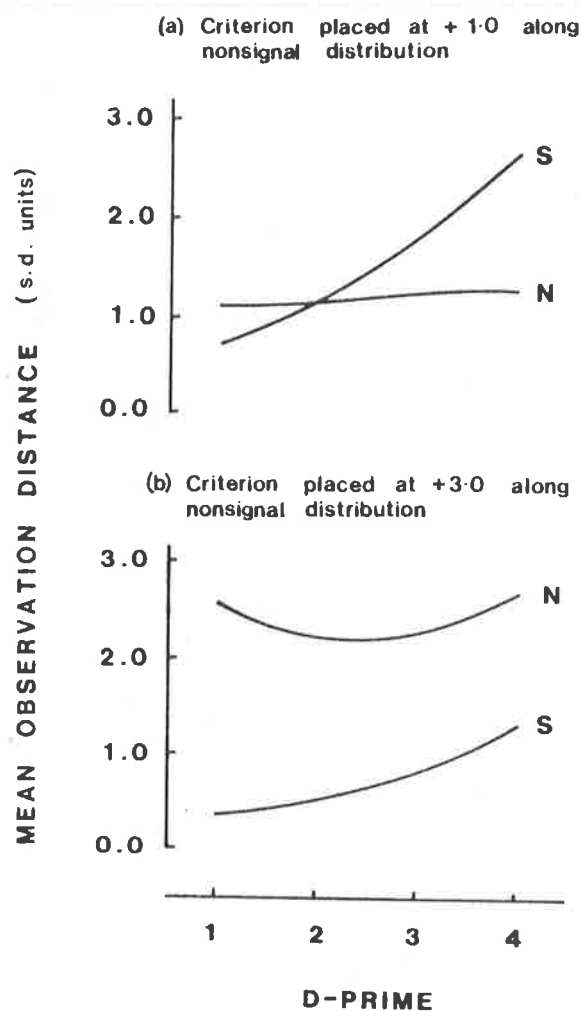


FIGURE 56 Mean observation distances of signal (S) and nonsignal (N) responses for values of d' between 1.0 and 4.0 when (a) the decision criterion is placed 1.0 standard deviation above the mean of the nonsignal distribution and (b) the criterion is placed 3.0 standard deviations above the mean of the nonsignal distribution.

In order to restore the previous balance of confidence between the two response, therefore, the decision criterion would need to be shifted upward. Such an adjustment might explain at least part of the increase in β between training and testing found by Baddeley and Colquhoun (1969). If detectability declines, rather than increases, the required criterion adjustment would be a shift downward, which thus agrees in general with the findings of Experiment 5.

The reason that there was no criterion adjustment over the initial stages of the decline in detectability in Experiment 5 could be due to the fact that confidence in both responses declines as detectability declines (Figure 56(a)), although to a much greater extent for signal responses. This effect will become more exaggerated as the value of the decision criterion is increased, as can be seen in Figure 56(b), where the data are for a criterion that is placed three standard deviation units above the mean of the nonsignal distribution. As d' in this instance falls from 4.0 there is first a period during which the mean observation distance of signals and nonsignals decline at comparable rates so that no criterion adjustment would be required. Then, as d' declines below 2.5, the mean distance for nonsignal responses increases while that of signal responses continues to decrease, so that downward adjustment of the criterion would be required.

Again, there could be a trade-off between the effect that the criterion adjustments have on restoring the balance of confidence in the two response alternatives, and the amount of effect that is tolerated on overall response confidence. Figure 57 gives the mean observation distance of signal and nonsignal responses, and their combined mean, at each level of detectability for the data of Experiment 5. These obtained values are compared with those which would have resulted if the observer had maintained the initial position of the decision criterion, that is, satisfied the Neyman-Pearson objective by holding false alarm rate constant. The empirical decline in the overall mean observation distance of all responses was greater than that which would have been associated with a fixed decision criterion. However, the criterion shifts which did occur again resulted in much less disparity between the mean observation distances of signal and nonsignal responses than would have occurred had no criterion adjustment taken place.

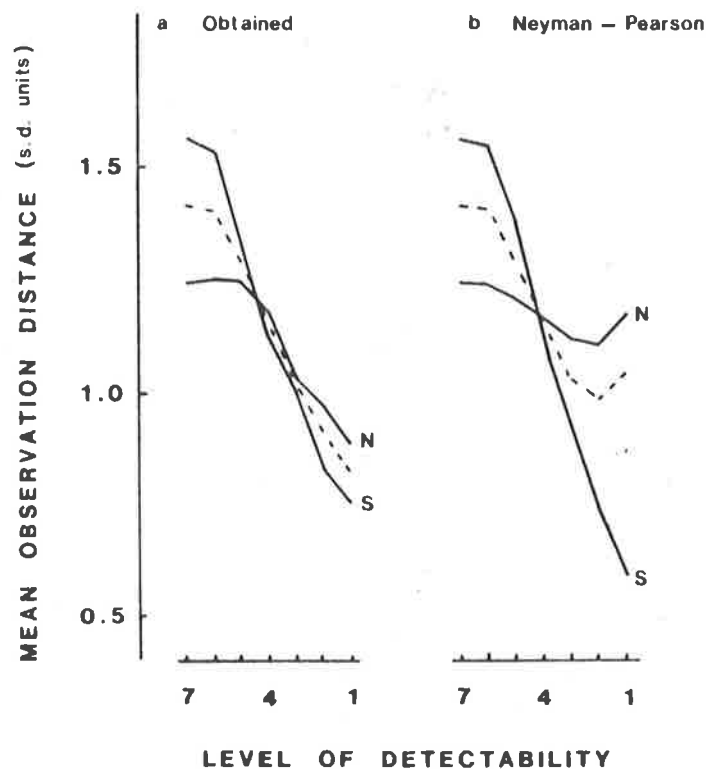


FIGURE 57 The mean observation distances of signal (S) and nonsignal (N) responses at each level of detectability in Experiment Five. The obtained distances (a) are compared with those which would have resulted if (b) the observer had retained the initial criterion position in order to satisfy the Neyman-Pearson objective. The combined average distance for all responses is given by the dashed line.

A final point to note about Experiment 5 is that if the observer had behaved in the ideal fashion specified by signal detection theory, the criterion would always have been placed at the intersection of the signal and nonsignal distributions. The mean observation distances of signal and nonsignal responses would then have been equal, and hence declined in the same manner, throughout the task. However, this would also have been the "ideal" result produced by the confidence-based model, if the observer had started the task with no differential response bias, rather than the bias that was observed toward making signal responses.

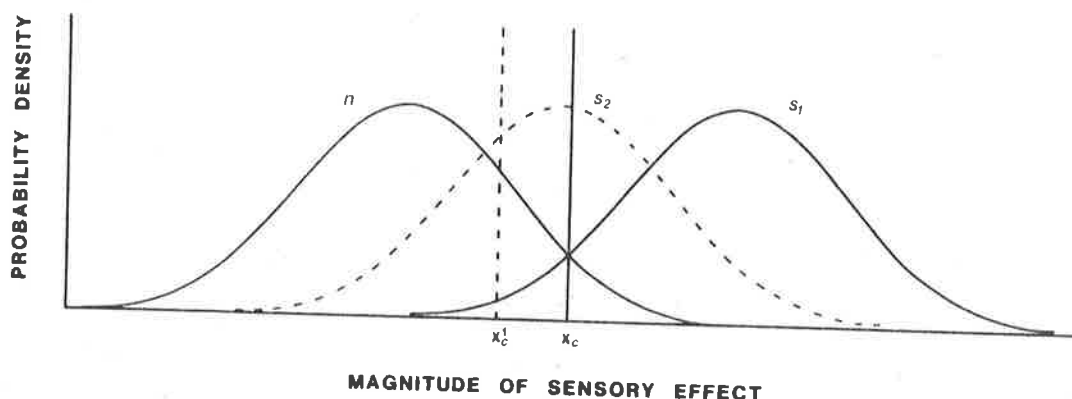


FIGURE 58 Hypothetical distributions of noise, n , a high intensity signal, s_1 , and a low intensity signal, s_2 , in the experiment of Gescheider, Wright and Polak (1971). With a decision criterion placed at x_c , the mean observation distance of signal responses relative to nonsignal responses will be greater when the probability of s_1 exceeds s_2 than when s_2 exceeds s_1 . A shift in the criterion position from x_c to x_c^1 when the probability of s_2 exceeds s_1 will increase the mean observation distance of signal responses relative to nonsignal responses.

Further empirical support for the model is contained in a study by Gescheider, Wright and Polak (1971) in which signals of various intensity were presented during each sequence of trials. The relative probability of being presented with signals of high or low intensity was varied between trial sequences, while the a priori probability of a signal was held constant. A theoretical simplification of this design is shown in Figure 58 where s_1 is the hypothetical distribution of a high intensity signal and s_2 that of a low intensity signal. When there is a higher probability of sampling an observation from s_1 rather than s_2 on signal trials, the mean observation distance of signal responses, relative to a criterion set at x_c , will be greater than when the probability of sampling an observation from s_2 is higher than s_1 . Thus, for an intermediate position of the criterion, such as x_c , when

the relative probability of s_2 increases, and the observer attempts to maintain the balance of confidence between signal and nonsignal responses, the criterion must be set at some lower position in order to increase the mean observation distance of signal responses relative to nonsignal responses. This is indicated in Figure 58 as a shift in the location of the criterion from x_c to x'_c . In agreement with this prediction, Gescheider et al. found that observers adopted a lower decision criterion when the relative probability of a low intensity signal was higher than that of a high intensity signal.

6. A Limitation of Response Bias as a Mediator in Feedback Control

Up to this point it has been assumed that whenever feedback about performance indicates that some corrective action must be taken, it is achieved through adjustment in the relative bias that the observer has toward making signal responses. Specifically, a single sensory cutoff, in the signal detection sense, is either raised or lowered in order to restore the confidence, or relative proportion, of a response to its desired value. Such a model is limited by the asymmetry of the effects that any adjustment of the decision criterion has upon the two response alternatives. This is most obvious when the response of a confidence-based mechanism to variation in signal detectability is considered. For instance, it was shown in the previous section that under certain conditions a decline in signal detectability will produce the same rates of decline in the mean observation distance of signal and nonsignal responses. As a result, the confidence of both responses cannot be increased toward target levels since no shift of the criterion can increase the mean observation distance of one response without decreasing that of the other. Even when considering changes in mean observation distances that proceed at different rates, this asymmetry will mean that a discrepancy

with target levels of confidence will necessarily exist for at least one of the responses, provided that the direction of change is the same for both responses. Hence, additional assumptions would have to be incorporated into the model in order for it to be applied in such circumstances.

Clearly, then, in order for a model incorporating response bias to effectively minimize discrepancy between the ongoing and reference values of a response variable which is under feedback control, changes in that variable must be asymmetric with regard to the two response alternatives. This limitation of the model becomes more crucial in the context of discrimination where, unlike signal detection, variation in discriminability can produce symmetrical effects on the two response alternatives that are not confined to a particular range of stimulus intensity, nor solely to the confidence response variable.

A standard type of discrimination task is that employed in Experiments 1 to 3, in which the observer had to judge whether line length was greater to the left or the right of a π -figure stimulus. The conceptualization of discrimination in terms of signal detection theory **is discussed in section B. 3. (b) of Chapter 3.** Briefly, on each trial the observer must decide whether the observed ($R - L$) difference in the sensory effect produced on the right, R , and on the left, L , was sampled from the distribution of differences that arise from presentations in which the right is objectively greater, r , or the distribution arising from presentations in which the left is objectively greater, l (see Figure 29). For each value of objective difference there is a pair of distributions, (l, r), which correspond to the effect produced by a stimulus in which the left is greater and the effect of one in which the right is greater. Shown in Figure 59 are two such pairs of hypothetical distributions, (l_1, r_1) arising from two stimuli of high discriminability, and (l_2, r_2) arising from two stimuli of low discriminability.

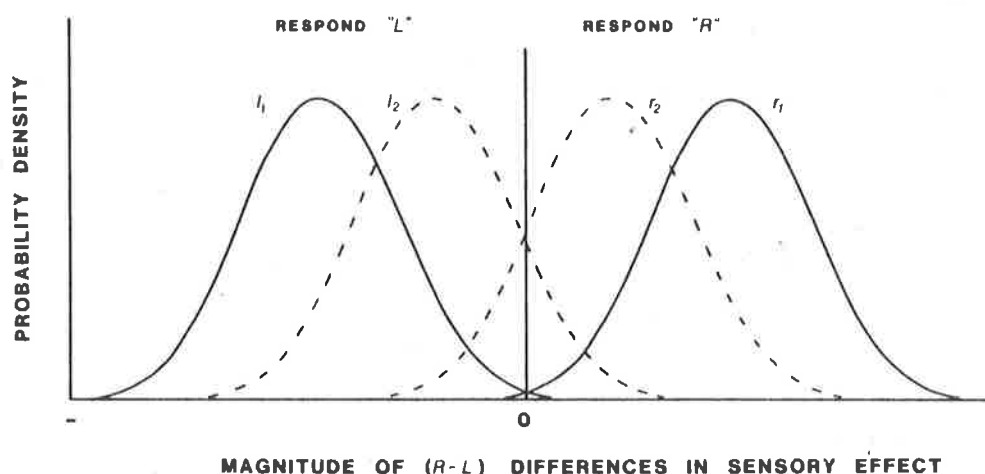


FIGURE 59 Hypothetical distributions of differences in sensory effect arising in a discrimination task in which either the left, L , or right, R , of two sources is of the greater magnitude. The distributions produced by one pair of stimuli (l_1, r_1) are the result of a greater objective difference between the two sources than exists between a second pair of stimuli (l_2, r_2).

For any particular pair of stimuli, if the two stimuli are presented with the same a priori probability and no biasing factors exist, most models predict that the observer would place the decision criterion at the zero-point along the sensory continuum as shown in Figure 59. An interesting situation then emerges if one considers the possible outcome of variation in discriminability during a sequence of trials. For instance, if a series of trials with the pair of stimuli corresponding to (l_1, r_1) were followed by a series with the pair associated with (l_2, r_2), a feedback model which controls performance through adjustments in response bias will be incapable of providing the observer with a means to counteract the effects on performance of the decline in stimulus discriminability. This is a consequence of the symmetrical effects which occur to "left" and "right" responses: as l_1

shifts to l_2 , and r_1 to r_2 , the probabilities of making each response correctly decline by the same amount and the mean observation distances of the two responses similarly decline to the same extent. Since the same changes occur to the conditional probabilities of each response, there is then no effect on unconditional response probabilities, and a model which uses response probability as feedback will have no cause to take action. Although a model which employs confidence would demand action due to the lowered mean observation distances, a shift of the criterion in order to increase the confidence of one response would decrease that of the other, so that no effective action could take place. A feedback mechanism based on response latencies would also face the same dilemma as one based on confidence.

It is conceivable, however, that the observer might make some form of adjustment, other than in response bias, in order to compensate for the decreases in response accuracy, latency and confidence which accompany a decline in stimulus discriminability. A means of detecting whether some form of adaptation does occur, given the above design, is to randomly intersperse a third pair of stimuli throughout the entire trial sequence, i.e. during the series of trials in which pair 1 is presented as well as the subsequent series in which pair 2 is presented. Since this additional pair of stimuli represents a constant stimulus difference, then, according to the response bias models, the performance of the observer when presented with these stimuli should not vary during the trial sequence. If, however, some form of adaptation occurs due to the decline in the general level of discriminability between those trials in which the additional pair of stimuli is presented with pair 1 and those in which it is presented with pair 2, then performance change can be expected to occur at this constant stimulus difference. The

following experiment was designed specifically to determine whether any such performance variation might be observed, and, if so, the sort of underlying process that it might suggest.

B. EXPERIMENT SIX : PILOT INVESTIGATION OF THE EFFECTS OF
A RAMP CHANGE IN DISCRIMINABILITY

The purpose of this experiment called for a task in which the general level of discriminability of stimuli varied, but, in order to detect whether the observer made any adjustment to this variation, stimuli of some constant level of discriminability occurred throughout the task. Further, a task was required in which there was minimal likelihood of bias toward either of the response alternatives. The task chosen was basically the same discrimination task as used in Experiments 1 to 3, with the stimuli modified in a way that allowed their discriminability to vary in the desired fashion.

1. Method

(a) Task and stimuli

On each trial observers were required to judge which was the longer of two vertical lines in a π -figure stimulus (see Figure 5) presented on the screen of a television monitor and viewed at a distance of 4.4 metres. The two lines were separated by a distance of 23 mm and the length of the longer line was held constant at 78 mm. The length of the shorter line varied between 73.9 mm and 77.5 mm in equal increments to create nine distinct "line length differences" labelled, in order of increasing magnitude, 1 to 9, which subtended visual angles of 0.006° to 0.054° . Five levels of discriminability were defined: the highest, level 5, comprised line length

differences 5 to 9, level 4 had differences 4 to 8, level 3 had differences 3 to 7, level 2 had differences 2 to 6 and the lowest, level 1, had differences 1 to 5. The mean line length difference of 5 thus occurred at all levels of discriminability and so allowed performance measurements at a constant stimulus difference.

(b) Apparatus

Stimuli were generated by a graphics program contained in a D.E.C. PDP 8/F computer. On each trial, the side on which the longer line occurred and the length of the shorter line was determined by a disk-based D.E.C. PDP 8/E computer. Stimuli were visible for 2 seconds unless terminated by a decision response. Following stimulus termination there was a further 2 second interval prior to the next stimulus presentation during which a confidence rating could be recorded, as well as the primary decision response if it had not been made while the stimulus was still present. The second computer classified the responses, measured the latency of the primary decision response and the value of the confidence response, and printed out summaries of these statistics after each block of training trials and again at the end of the main session.

The right or left of two perspex buttons, lying on either side of a wedge-shaped resting position, was depressed with the index finger of the dominant hand to indicate whether the right or left line, respectively, was judged to be longer. Confidence was then indicated by depressing one of a row of ten buttons, divided in the middle by a yellow line so that five lay in front of each primary response button. For both responses the confidence scale ran from "least confident" at the innermost button lying next to the yellow line, to "most confident" at the outermost button.

(c) Design

An experimental session contained two blocks of 50 training trials each, followed by 1000 test trials. There was a three minute interval after each block of training trials during which a print-out of performance was examined. The experimental sequence, which could last for a maximum of 67 minutes, was divided into 20 blocks of 50 trials each and presented as an unbroken series of trials. During each training and test block stimuli were presented at each of the five line length differences for that particular discriminability level. The longer line occurred to the left and right in random order with the constraint that an equal number be presented on each side at each line length difference within each trial block.

The discriminability level of 5 was employed during the training trials in order to build up decision speed prior to the main session. The results obtained from three pilot subjects indicated fairly low error rates at all levels of discriminability so that no training at the lower levels was considered necessary. The test sequence was divided into five main periods of 350, 100, 100, 100 and 350 trials, respectively. As shown in Figure 60, the level of discriminability employed during each period was 1, 2, 3, 4, 5, respectively, among one group of observers (condition INC) and 5, 4, 3, 2, 1, respectively, among the other group (condition DEC). Therefore, in condition DEC the level of discriminability was progressively lowered in a ramp fashion during the experimental sequence, while, in condition INC, it was progressively increased.

Two experimental conditions were employed, differing in the direction in which the ramp change in discriminability proceeded, in order to counter-balance what appeared to be an effect of learning early in the task, and an effect similar to fatigue later in the task, among the first 14 observers

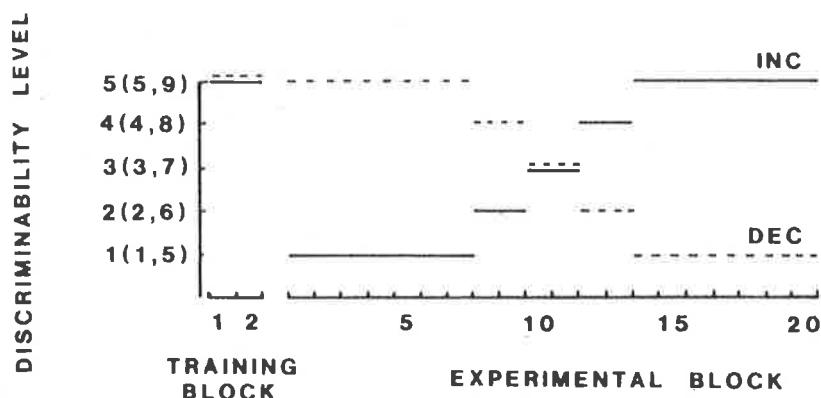


FIGURE 60 The basic design of Experiment 6. The level of discriminability at which stimuli were presented was increased during the task for the INC group of observers and decreased for the DEC group of observers. The range of line length differences at each discriminability level is given in brackets by the number for that level.

who were tested. These observers were tested under condition DEC and displayed a tendency for performance to improve across discriminability level 5, at the start of the session, and then decline across level 7, at the end. Fourteen additional subjects were therefore tested with the levels of discriminability presented in ascending order.

(d) Procedure

Observers were tested individually and in complete isolation. Watches were removed upon entering the experimental chamber. Prior to the first block of training trials instructions were given concerning the nature of the discrimination task and the use of the response buttons and confidence scale. The need to make both responses within four seconds of stimulus onset was stressed. Observers were told that the difficulty of the discriminations would vary during the main session to follow, but not told the exact nature of this variation. They were then told that practice would be at the lowest level of difficulty and to try to spread their confidence

ratings over most of the scale, even though they might feel fairly confident in the majority of their decisions. Prior to the main sequence, further instructions were given, which emphasized that each confidence rating was to be based on the immediate past decision and not on any more general feeling of confidence. In particular, observers were told not to be concerned if it seemed at any time that the majority of confidence ratings were being made at either extreme of the confidence scale (as might have been the case during the practice trials).

(e) Observers

Twenty-eight males enrolled in first year Psychology at the University of Adelaide volunteered to participate in the experiment as part of their course requirements. Fourteen were tested under each experimental condition. Ages ranged from 17 to 27 years, with a median age of 19 years. All observers had normal or corrected vision.

2. Results

(a) Pre-test differences between experimental groups

In view of the finding that the differences in performance between the two experimental groups of Experiment 5 had existed prior to the main experimental session, and the fact that observers had not been randomly assigned to treatment conditions in the present experiment, tests were carried out to determine whether any pre-test differences existed between the two experimental groups. Analyses of variance were performed on the data from the second block of practice trials and the results are given in Table 24. Only statistics associated with correct responses are reported due to the extremely low incidence of errors among both groups of observers.

TABLE 24. Results of analyses of variance on the data from the second block of practice trials. Response measures were averaged across the five line length differences.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Correct right response probability</u>				
	Condition (C)	1	.060	6.720	.05
	S within C	24	.008		
(b)	<u>Correct right response latency</u>				
	Condition (C)	1	.599	3.565	
	S within C	24	.168		
(c)	<u>Correct right response confidence</u>				
	Condition (C)	1	30.627	12.072	.005
	S within C	24	2.536		
(d)	<u>Correct left response probability</u>				
	Condition (C)	1	.004	.923	
	S within C	24	.005		
(e)	<u>Correct left response latency</u>				
	Condition (C)	1	.674	5.819	.05
	S within C	24	.115		
(f)	<u>Correct left response confidence</u>				
	Condition (C)	1	11.760	7.406	.05
	S within C	24	1.587		

The significant effects of experimental condition on the probability and confidence of correct right responses were due to the higher probability (1.0 vs. .96) and confidence (4.4 vs. 3.4) of these responses among the observers in condition INC. Similarly, the effects on the latency and confidence of correct left responses reflected the lower latency (865 msec vs. 1009 msec) and higher confidence (4.5 vs. 3.9) of these responses among the same observers. These differences indicate that performance prior to the main experimental

sequence was higher among observers in condition INC than it was among observers in condition DEC. Observers had been randomly chosen from the same student population, but conditions were run in succession. Although great care was exercised to maintain stability of experimental conditions, the possibility cannot be excluded that the differences between groups may have arisen, at least in part, from a change in the calibration of the television monitor.

Despite the fact that the two groups differed prior to the main trial sequence, experimental condition was still retained as a factor in the analysis of data. Although any group differences in the overall level of performance could not be attributed to the manner in which discriminability was varied, it was felt that an interaction between experimental condition and discriminability level might suggest some effect of the direction in which the ramp change in discriminability took place. Subsequent data analyses then were aimed mainly at determining, firstly, whether there was an effect of discriminability level on performance measures averaged across all line length differences at each level of discriminability, and secondly, whether there was any effect of discriminability level on performance at the constant line length difference (5). The former analysis was necessary to determine whether the observer's performance was sensitive to change in the range of discriminability employed, so that any results which emerged in the latter analysis could be attributed to this variation. Data from the first and last five blocks of 50 trials were excluded from analysis in order to equalize the total number of experimental trials (100) which entered analysis from each level of discriminability, and to further eliminate any extraneous effects of practice or fatigue.

(b) Effects of variation in discriminability on overall performance

This section reports the results of analyses performed on performance measures averaged across all line length differences at each level of discriminability.

(i) Response probability. The results of analyses of variance on the conditional probabilities of correctly reporting the longer line when it occurred on the right and on the left are given in Table 25. The significant effect of condition on correct response probability is evident in Figure 61. The observers who were presented with a ramp increase in discriminability (condition INC) had the higher probability of a correct right response at all discriminability levels, a result consistent with that found in training. The main effect of discriminability level and its interaction with condition on both conditional probabilities can be seen in Figure 61 as a tendency for the probability of a correct response to be directly related to discrim-

TABLE 25. Results of analyses of variance on response probabilities averaged across the line length differences at each level of discriminability.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	Correct right response $P(R/r)$				
	Condition (C)	1	2.082	10.640	.005
	S within C	26	.195		
	Discriminability (D)	4	1.485	64.023	.00001
	C x D	4	.163	7.044	.00005
	D x S within C	104	.023		
(b)	Correct left response $P(L/l)$				
	Condition (C)	1	.028	1.033	
	S within C	26	.027		
	Discriminability (D)	4	.051	11.429	.00001
	C x D	4	.018	4.076	.005
	D x S within C	104	.004		

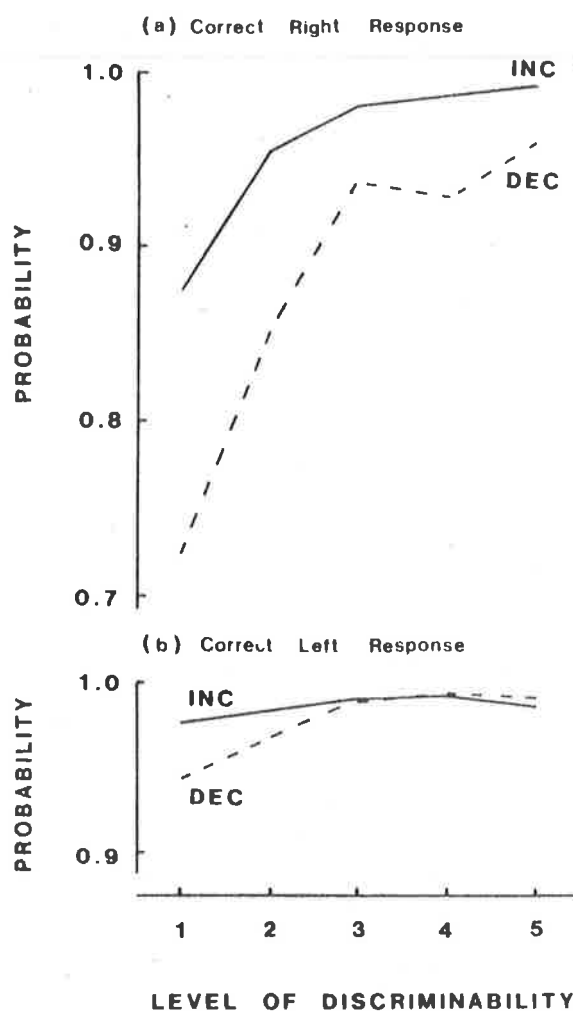


FIGURE 61 The mean conditional probability of making (a) a correct right and (b) a correct left response across the line length differences at each level of discriminability for each group of observers.

inability level among both groups of observers, the degree to which it varied with discriminability level being greatest among those observers presented with a ramp decline in discriminability.

Figure 61 indicates that among both groups of observers the probability of making a correct decision was asymptotic across the higher discriminability levels. This was confirmed by analyses of variance for the effect of discriminability level across levels 3 to 5 ($F = 3.048$, $d.f. = 2/52$, $p > .05$ for right responses; $F = .831$, $d.f. = 2/52$, $p > .4$ for left responses), and could explain

the interaction effects on both conditional probabilities, and lack of an effect due to condition on correct left response probability. In condition INC, observers began the task close to the performance 'ceiling', that is with conditional correct response probabilities close to 1.0. The extent to which the performance of this group could improve as discriminability increased from its lowest level was then less than the extent to which it could decline among the other observers as discriminability declined toward the same level. Hence, across the three highest levels of discriminability there was no interaction between condition and discriminability level on either correct right or correct left response probability ($F < 1.0$ in both cases) and performance appeared asymptotic. However, across levels 1 to 3 there were interactions on both correct right and correct left response probability ($F = 8.150$, d.f. = $2/52$, $p > .001$, and $F = 3.292$, d.f. = $2/52$, $p < .05$, respectively), indicating that the decline in response accuracy from its asymptote among observers in condition DEC was greater than the increase toward asymptote in condition INC. As can also be seen in Figure 61, correct response probability had a much greater range of values for right responses than for left, the probability of a correct left response being very high throughout the task. This narrower range, coupled with invariant performance over the higher levels of discriminability, resulted in no significant difference between the two groups of observers.

(ii) Response latency. Due to the high probability of making a correct decision at all levels of discriminability, no reliable statistics could be performed on the latencies of the two types of incorrect response. Table 26 gives the results of analyses of variance carried out on the latencies of the two types of correct response, as well as overall latency (the mean latency of all responses, whether correct or incorrect).

TABLE 26. Results of analyses of variance on response latencies averaged across all line length differences at each level of discriminability.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Correct right response</u>				
	Condition (C)	1	8.614	4.350	.05
	S within C	26	1.980		
	Discriminability (D)	4	2.359	32.339	.00001
	C x D	4	.087	1.203	
	D x S within C	104	.072		
(b)	<u>Correct left response</u>				
	Condition (C)	1	5.971	4.087	
	S within C	26	1.460		
	Discriminability (D)	4	1.189	31.587	.00001
	C x D	4	.071	1.911	
	D x S within C	104	.037		
(c)	<u>Overall</u>				
	Condition (C)	1	7.835	4.939	.05
	S within C	26	1.586		
	Discriminability (D)	4	1.825	38.177	.00001
	C x D	4	.059	1.249	
	D x S within C	104	.047		

The significant effects reported in Table 26 are readily apparent in Figure 62. Latencies of the two types of correct response, as well as overall latency, were inversely related to the level of discriminability. The significant effect of condition on correct right and overall response latency, and the marginal effect on correct left response latency (the F-ratio in Table 26 has $p = .053$), are shown in Figure 62 as consistently lower latencies among the observers in condition INC (again consistent with the finding from practice trials). The absence of any interaction indicates that the superior performance of this group was exhibited to the same extent at all levels of discriminability. Thus, unlike the case with response probabilities, the absence of a performance ceiling on response times, allowed for greater variation in the measure. As a result, more comparable ranges of values were shown by the two groups across the five levels of discriminability.

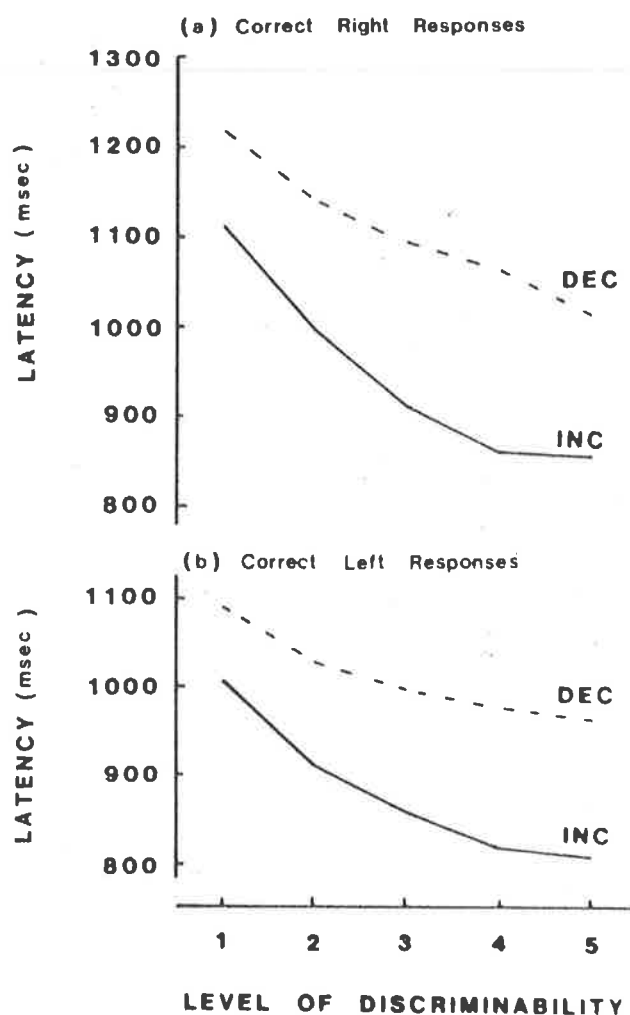


FIGURE 62 The mean latencies of (a) correct right and (b) correct left responses across the line length differences at each level of discriminability for each group of observers.

(iii) Response confidence. As was the case with response latencies, no reliable analysis could be performed on the confidence associated with incorrect responses. The results of analyses of variance on confidence in correct right and correct left responses, as well as the overall mean confidence expressed in all responses, are reported in Table 27. The significant main effects of condition and discriminability level on all three measures can be seen in Figure 63. Confidence was directly related to the level of discriminability and was highest among observers in condition INC at all

TABLE 27. Results of analyses of variance on response confidence averaged across all line length differences at each level of discriminability.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Correct right response</u>				
	Condition (C)	1	661.663	24.514	.00005
	S within C	26	26.990		
	Discriminability (D)	4	27.877	33.174	.00001
	C x D	4	2.214	2 .635	.05
	D x S within C	104	.840		
(b)	<u>Correct left response</u>				
	Condition (C)	1	407.700	20.213	.0005
	S within C	26	20.169		
	Discriminability (D)	4	24.907	29.729	.00001
	C x D	4	1.105	1.319	
	D x S within C	104	.837		
(c)	<u>Overall</u>				
	Condition (C)	1	528.960	24.225	.00005
	S within C	26	21.835		
	Discriminability (D)	4	26.963	40.658	.00001
	C x D	4	1.264	1.906	
	D x S within C	104	.663		

levels of discriminability. This latter finding is in agreement with the trends shown in response latency, and is also consistent with the differences found in practice.

The interaction between condition and discriminability level on correct right response confidence is not immediately apparent in Figure 63. However, it appears to be solely due to a greater increase in confidence among observers in condition INC, as discriminability increased from level 1 to 2, than to a decrease in confidence among observers in condition DEC, as discriminability dropped from level 2 to level 1. Thus there was a significant interaction when confidence was compared across levels 1 and 2 ($F = 21.519$, d.f. 1/26, $p < .0001$), but no other comparison of successive discriminability levels produced an interaction approaching significance ($F < 1.0$ in all cases).

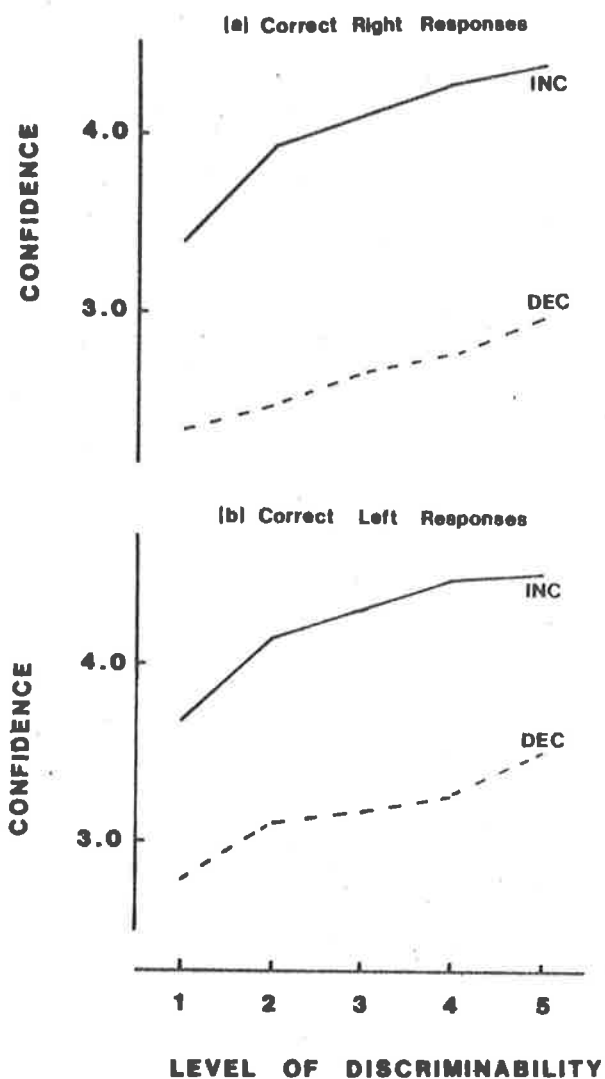


FIGURE 63 The mean confidence expressed in (a) correct right and (b) correct left responses across the line length differences at each level of discriminability for each group of observers.

As with response latencies, the two groups of observers tended to differ by the same extent throughout the task.

(c) Performance at the constant stimulus difference

The results of the analysis of probability, latency, and confidence measures, averaged across the line length differences at each level of discriminability, indicated that the performance of both groups of observers was sensitive to variation within the range of discriminability employed. In order to determine whether this variation had produced any sort of adaptation by the observer, only the data obtained at the line length difference of 5 entered the analyses to be reported in the present section, which is thus concerned with responses made to a stimulus difference that remained constant throughout the entire trial sequence.

(i) Response probability. Table 28 gives the results of analyses of variance on the conditional probabilities of making a correct response. The only significant effect was of experimental condition on the probability of a correct right response. As can be seen in Figure 64, observers in condition INC had the higher probability at all levels of discriminability,

TABLE 28. Results of analyses of variance on response probabilities at line length difference 5.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	Correct right response $P(R/r)$				
	Condition (C)	1	.165	5.077	.05
	S within C	26	.032		
	Discriminability (D)	4	.006	1.083	
	C x D	4	.008	1.334	
	D x S within C	104	.006		
(b)	Correct left response $P(L/l)$				
	Condition (C)	1	.003	.588	
	S within C	26	.005		
	Discriminability (D)	4	.002	1.049	
	C x D	4	.003	1.361	
	D x S within C	104	.002		

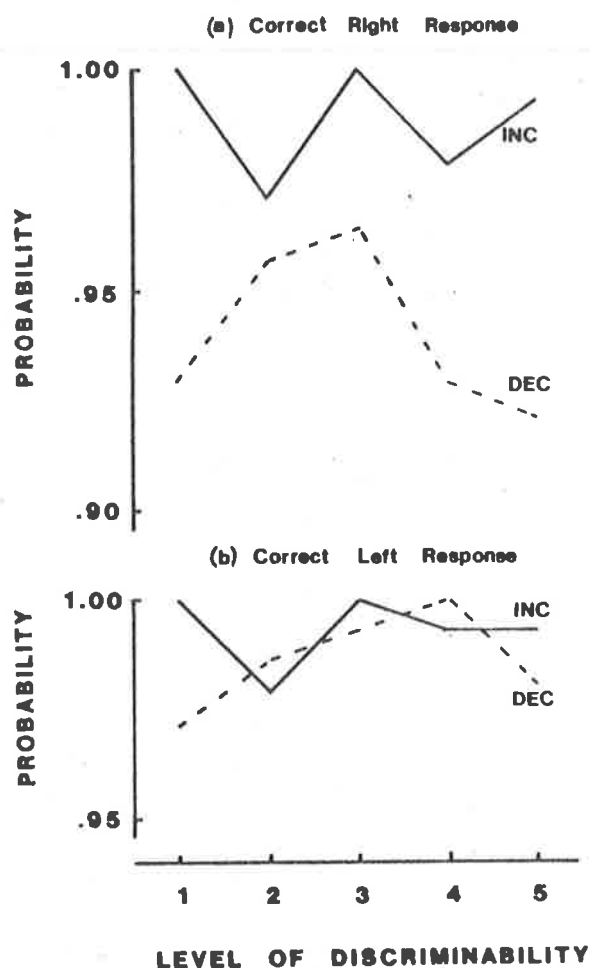


FIGURE 64 The mean conditional probability of making (a) a correct right and (b) a correct left response to stimuli with a line length difference of 5 at each level of discriminability for each group of observers.

consistent with their performance averaged across all line length differences. There were no significant effects on the probability of making a correct left response, which is not surprising since the accuracy of decisions among both groups when the longer line was presented to the left was almost at its maximum at the particular amount of stimulus difference associated with line length 5 (see Figure 64(b)).

(ii) Response latency. The infrequency of incorrect responses again prevented any reliable analysis of their associated latencies or confidence ratings. Table 29 gives the results of analyses of variance on the latencies of correct left and right responses, as well as the mean latency of all responses. The main effect of condition on all three measures can be seen in Figure 65 as consistently lower latencies among the observers in condition INC. The lack of any interaction with discriminability level again indicates that the degree to which response latencies were shorter among these observers remained constant. The significant effect of discriminability level on the latency of correct left responses can be seen in Figure 65(b). There was an overall tendency for these latencies to be inversely related to the level of discriminability. The effect appears to have been strongest,

TABLE 29. Results of analyses of variance on response latencies at line length difference 5.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Correct right response</u>				
	Condition (C)	1	2.131	5.170	.05
	S within C	26	.412		
	Discriminability (D)	4	.003	.177	
	C x D	4	.037	1.898	
	D x S within C	104	.019		
(b)	<u>Correct left response</u>				
	Condition (C)	1	1.231	4.577	.05
	S within C	26	.269		
	Discriminability (D)	4	.037	4.149	.005
	C x D	4	.011	1.242	
	D x S within C	104	.009		
(c)	<u>Overall</u>				
	Condition (C)	1	1.770	5.551	.05
	S within C	26	.318		
	Discriminability (D)	4	.015	1.545	
	C x D	4	.003	.334	
	D x S within C	104	.010		

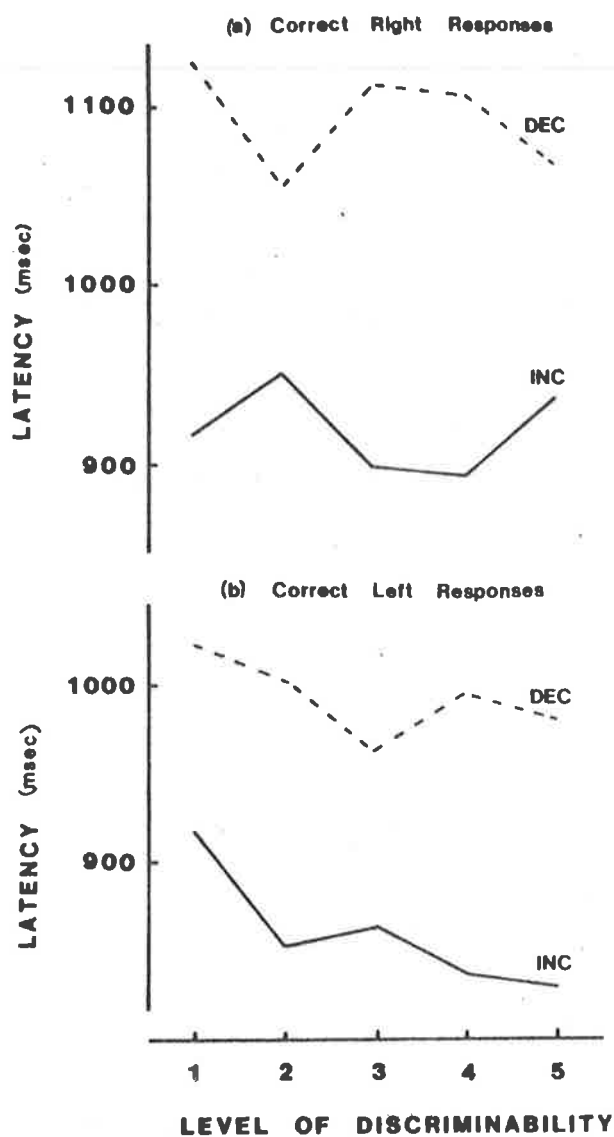


FIGURE 65 The mean latencies of (a) correct right and (b) correct left responses to stimuli with a line length difference of 5 at each level of discriminability for each group of observers.

when the combined performance of both groups is considered, across the lower levels of discriminability. Hence, an analysis of variance incorporating only the data from discriminability levels 3 to 5 yielded no significant effect of discriminability level.

(iii) Response confidence. The results of analyses of variance on the confidence expressed in correct right and left responses, as well as the mean confidence expressed in all responses, are given in Table 30. The significant main effect of condition on each of the measures is apparent in Figure 66 as a higher degree of confidence in decisions among observers in condition INC. The lack of any effect of discriminability, or its interaction with condition, is evidenced by nearly flat and parallel functions for the confidence measures.

TABLE 30. Results of analyses of variance on response confidence at line length difference 5.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Correct right response</u>				
	Condition (C)	1	165.242	31.588	.00001
	S within C	26	5.231		
	Discriminability (D)	4	.116	.439	
	C x D	4	.134	.505	
	D x S within C	104	.265		
(b)	<u>Correct left response</u>				
	Condition (C)	1	91.657	21.817	.00001
	S within C	26	4.201		
	Discriminability (D)	4	.424	1.708	
	C x D	4	.217	.910	
	D x S within C	104	.238		
(c)	<u>Overall</u>				
	Condition (C)	1	126.106	27.547	.00005
	S within C	26	4.577		
	Discriminability (D)	4	.027	.160	
	C x D	4	.138	.805	
	D x S within C	104	.171		

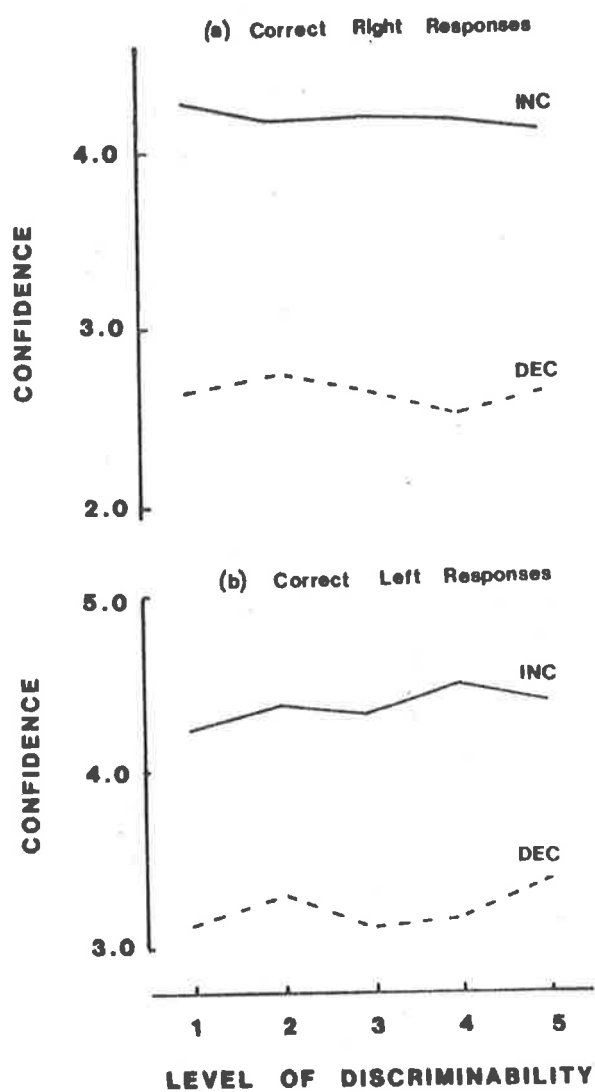


FIGURE 66

The mean confidence expressed in (a) correct right and (b) correct left responses to stimuli with a line length difference of 5 at each level of discriminability for each group of observers.

3. Discussion

(a) Sensitivity of overall performance to the variation in discriminability

The ramp change in discriminability produced consistent trends in performance measures averaged across the line length differences at each level of discriminability. The probability and confidence of correct decisions were directly related, and their latency inversely related, to the level of discriminability. These relationships were in agreement with the distance-from-criterion hypothesis of signal detection: a decline in the average line length difference would be associated with sensory distributions that lie progressively closer to the decision criterion (Figure 59). Thus the proportion of observations sampled from the appropriate side of the decision criterion would decrease, as would also the mean distance of these observations from the criterion so that their consequent responses would decline in speed and confidence.

As was expected, the effects produced on the two response alternatives were symmetrical, despite a seeming bias toward making the left response, as indicated by the relatively higher probability of making that response correctly. Such high accuracy was also shown in right responses across the higher discriminability levels, so that a performance asymptote was reached in response accuracy that did not allow the same degree of variation across discriminability levels that was displayed by response latency and confidence. Lastly, no differential effects were found to result from the direction in which the ramp change in discriminability proceeded, other than could be ascribed to between-group differences that existed prior to the experimental session.

(b) Evidence of adaptation to variation in discriminability

There was scant evidence of performance change at the constant line length difference as discriminability varied, the only significant result being an inverse relationship between discriminability level and the latency of correct left responses. Even though this means that out of six possible cases only one result indicated any sort of adaptation during the variation in discriminability, the finding should not be dismissed offhand. Firstly, the particular amount of stimulus difference chosen to test for performance adaptation elicited very high response accuracy so that there was a very narrow range within which variation in that measure could take place. Hence, the latency measure may have been the more sensitive index of any adjustment demanded by the changes in overall performance during the ramp change in discriminability. Secondly, the finding cannot have resulted from the influence of some extraneous variable such as fatigue. Such an effect would have proceeded in the same direction among both groups of observers during the course of the task. Thus, the direction of the relationship between latency and discriminability would have been reversed in observers who experienced a ramp increase in discriminability, when compared to those who experienced a ramp decrease. However, the same sort of relationship was shown by both groups. Lastly, although no effect was shown in response confidence, the design of the task itself may have made it unlikely that an appreciable change would have been found in any measure. While changes in overall performance were consistent across the levels of discriminability, they were of relatively small magnitude, especially across the higher levels of discriminability (see Figures 61 to 63). Thus, if an observer monitored changes in, for instance, response confidence in order to determine when adjustment was necessary, variations in that response attribute may have

been imperceptible during most of the task. Therefore, the gradual manner in which discriminability varied, allied with a generally high standard of performance, to a great extent may have rendered the task powerless to produce many measurable instances of the type of adaptive behaviour which was sought.

(c) The possibility that *caution* had been adjusted

If the tentative assumption is made that the effect found in response latency indeed represents an adjustment to the variation in discriminability, a serious problem would exist for all the previous models which have used response bias to mediate in a feedback process. This follows, since any relationship between left response latency and discriminability level should have caused the opposite relationship to be shown between right response latency and discriminability. For instance, if the result had been due to a shift away from the means of the l distributions toward the r distributions as discriminability declined (Figure 59), then the increase in the latency of left responses should have been accompanied by a decrease in that of right responses. The result, however, is compatible with a change in caution, rather than bias, according to a data accumulation model in which there was also a tendency for bias toward the left response, as indicated by the conditional response probabilities of the present study (Figure 61). The basic properties of data accumulation models will be briefly discussed in order to clarify this explanation.

(i) Data accumulation models and caution. Data accumulation models differ markedly from the signal detection model which has been utilized thus far in that, rather than make a decision based on a single observation of sensory difference, the observer is assumed to make a series of observations

during each observation interval (two models of this sort were mentioned in section C. 3. (a). (iii) of Chapter 3). Observations are made sequentially, and at a steady rate, each one providing evidence in favour of one of the decision alternatives. That is, to make an analogy with the present investigation, each observation would be of either a positive ($R-L$) sensory difference, favouring the right response, or a negative difference favouring the left response. During the course of the observation interval these differences will vary according to which sensory distribution is presently being sampled, and these distributions can again be assumed to vary over time in the manner shown in Figures 29(b) and 59.

Each type of data accumulation model specifies some "optional-stopping" decision rule according to which observations continue until some criterion C is satisfied. For instance, in a 'runs' model (Audley, 1960) observations continue until C consecutive positive or negative differences have been observed. In a 'random walk' model (Stone, 1960), observations continue until there is a difference of C between the total of one difference and that of the opposite sign. In a 'recruitment' model (La Berge, 1962), observations continue until there is a total of C positive or negative differences. Finally, in an 'accumulator' model (Vickers, 1970), observations continue until the magnitude of positive or negative differences that have been sampled exceed some total C . Thus C can be a critical number of observations of sensory difference, as in the first three models, or a critical magnitude of observed stimulus difference, as in the latter model. Since, according to all the models, observations are made at a steady rate, mean reaction time will be directly related to the value of C .

A distinction, both in terminology and function, can be made between response bias and response caution that is not possible in the signal detection

model. To use a recruitment process as an example, response bias would refer to unequal critical totals for the two responses, such as the situation shown in Figure 67(b) where there is a bias toward making the left response. Here the critical total of negative ($R-L$) differences, C^- , required in order to make a left response, is less than that of positive differences, C^+ , required for a right response. At a given level of discriminability, the average time taken to make left responses will then be less than that of right responses. Caution, on the other hand, would refer to the absolute, rather than relative, sizes of the critical totals, so that an increase in caution would be associated

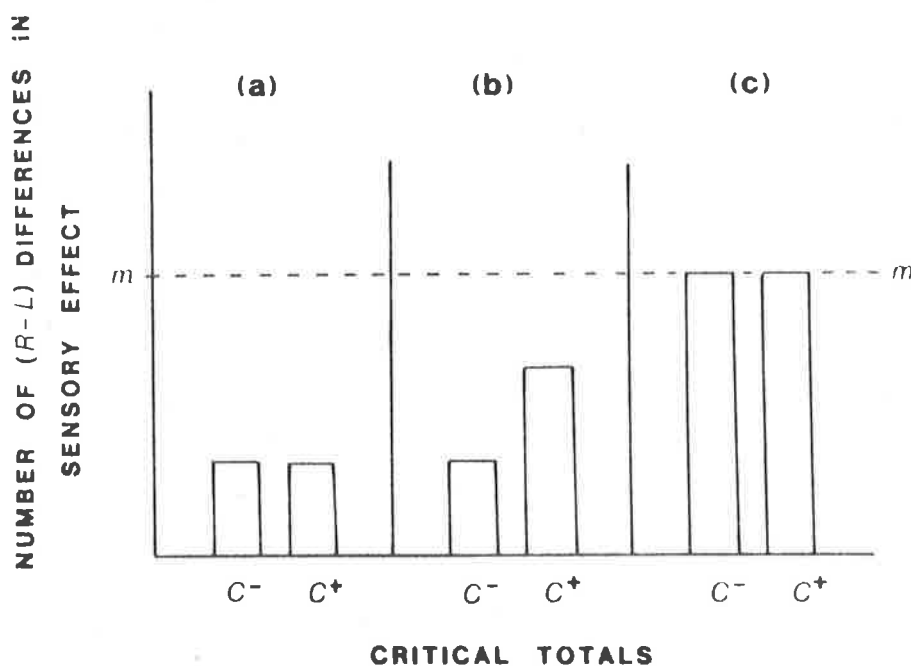


FIGURE 67

Hypothetical critical totals, C^+ and C^- , required in order to reach a respective right or left decision. In (a) there are low unbiased critical totals. In (b) there is bias toward the left response. In (c) the critical totals have been raised to a value, m , above which the time taken to reach either one of the critical totals could exceed the length of the observation interval.

with increases in both C^+ and C^- . This is shown between Figures 67(a) and 67(c), where the critical totals required for both responses have been increased, resulting in a general increase in response times. The signal detection model so far considered cannot produce such a change in caution, since any adjustment to the decision criterion will lower the latency of one response and raise that of the other.

(ii) The present experimental effect. In the present experiment, conditional response probabilities indicated that there was a bias toward making the left response. Then, as discriminability declined, there was an increase in the time taken to make that response to a constant stimulus difference, but no effect on the latency of a right response to the same stimulus difference. According to a recruitment process (and, indeed, any of the data accumulation models considered above), this is consistent with a reduction in the bias toward making the left response through an increase in C^- . However, since the characteristics of both responses were affected in the same manner by the decline in discriminability, any adjustment on the part of the observer that was intended to offset the general effects of this decline should have had the same effect on both responses. Thus, if the observer responds by increasing criterion level, it should be increased for both responses; caution, not bias, would then be adjusted. It is possible, however, that bias toward the left response had obscured the effect of an increase in caution on the latency of right responses. Since there is a limit to the number of observations which can be made during the interval in which the stimulus is visible, there is probably an upper limit placed on the values that the critical totals can take; otherwise a decision might not be reached during the observation interval. If we assume this upper

limit to be the same for all responses, then, if response bias exists, a subsequent increase in caution could produce the effect on the critical totals that is shown between Figures 67(b) and 67(c). The amount to which the critical total of the favoured response can increase is greater than that of the non-favoured response. In this way, the increase in the latency of left responses in the present study might have reflected the greater extent to which the criterion for a left response, rather than a right response, could increase as caution was increased. Conversely, a decline in caution would allow a situation in which the bias toward left responses could be displayed.

(iii) The need to confirm the effect. The last assumption, an unbiased upper limit on the critical totals, may seem rather premature to account for an effect as being due to a change in response caution, especially since evidence of such an effect appeared only in one response measure. However, if level of caution can be shown conclusively to be affected by variation in discriminability, then specific patterns of behaviour are predicted by the 'adaptive accumulator' model of Vickers (1978, 1979), which, unlike the other data accumulation models mentioned, incorporates the principle of self-regulation. Further, these predictions can be directly contrasted with those made by a signal detection process, altered in such a way that it incorporates response caution as well as bias, and employs a similar goal in the feedback process to that of the adaptive accumulator model.

As has been noted, the design of the present experiment may have produced adjustments that were too infrequent or indistinct to be generally measurable. For this reason, it was decided to employ a further task, in which the basic discrimination required was more complex than the present

one, and in which variation in discriminability was more abrupt than the present ramp change. By this means, it was hoped that performance would be less accurate, and the effects of any contingent variation in caution, or other possible mediator in a feedback process, would be enhanced.

CHAPTER 6

A. CRITERION REGULATION AND GENERAL CAUTION IN RESPONDING

The previous exploratory experiment suggested that the observer might respond to a decline in discriminability by adopting more cautious behaviour. The possibility that an observer will make slower and more deliberate responses when the demands of a task increase has much intuitive appeal, but as yet remains an ad hoc suggestion. However, such compensatory behaviour will be expected to follow from the operation of the 'adaptive accumulator' mechanism that has been advanced by Vickers (1978, 1979), in which the characteristics of a primary decision process are adjusted when there is variation in the statistical characteristics of sensory input during a series of trials. Variations in either the discriminability or probability of stimuli will dictate the required adjustment in either bias or caution. The general features of the model will now be presented, followed by those of a signal detection model that also incorporates response caution, and is based on similar principles of self-regulation.

1. The Adaptive Accumulator Model

Vickers (1978, 1979) has developed a general-purpose decision module which has the capacity to operate at any level of information processing. In his model of perception, this mechanism operates at the primary level to classify sensory events according to the available response alternatives, while, at the secondary level, a similar mechanism assesses the confidence arising from primary decisions in order to determine if adjustments are necessary in the primary decision process. Although structurally equivalent, the two processes will be considered separately.

(a) The primary decision process

The general assumptions upon which the primary decision process are based were presented in the discussion of data accumulation models in the previous chapter (section B. 3. (b) (i)). During each stimulus presentation, the observer makes a series of observations, at a steady rate, each of an $(R - L)$ stimulus difference in order to determine whether the sample of differences arose from the distribution, r , associated with a stimulus in which R is objectively greater than L , or the distribution, l , associated with a stimulus in which R is objectively less than L . The process by which a decision is reached is shown in Figure 68. Positive and negative differences are accumulated separately, the magnitude of each difference being added to a total, t_g , if it is positive, or a total, t_l , if it is negative, until a critical total, k_g or k_l , respectively, has been reached, whereupon the observer initiates the "greater" ($R > L$) or "lesser" ($R < L$) response. The general three-category process is depicted in Figure 68, in which an "equals" ($R = L$) response is triggered whenever the sums of the totals, t_g and t_l , minus the absolute value of their difference reaches a critical total, k_e , before either k_g or k_l has been reached.

The model can be applied to a number of experimental paradigms either by suppressing responses through adopting high critical totals, or by varying the connections between the primary decisions and the allowable response alternatives. For example, in situations where only the two extreme responses are allowed, such as in the previous experiment, it is assumed that a very high value is assigned to k_e in order to make an "equals" decision unlikely and thus suppress the intermediate response.

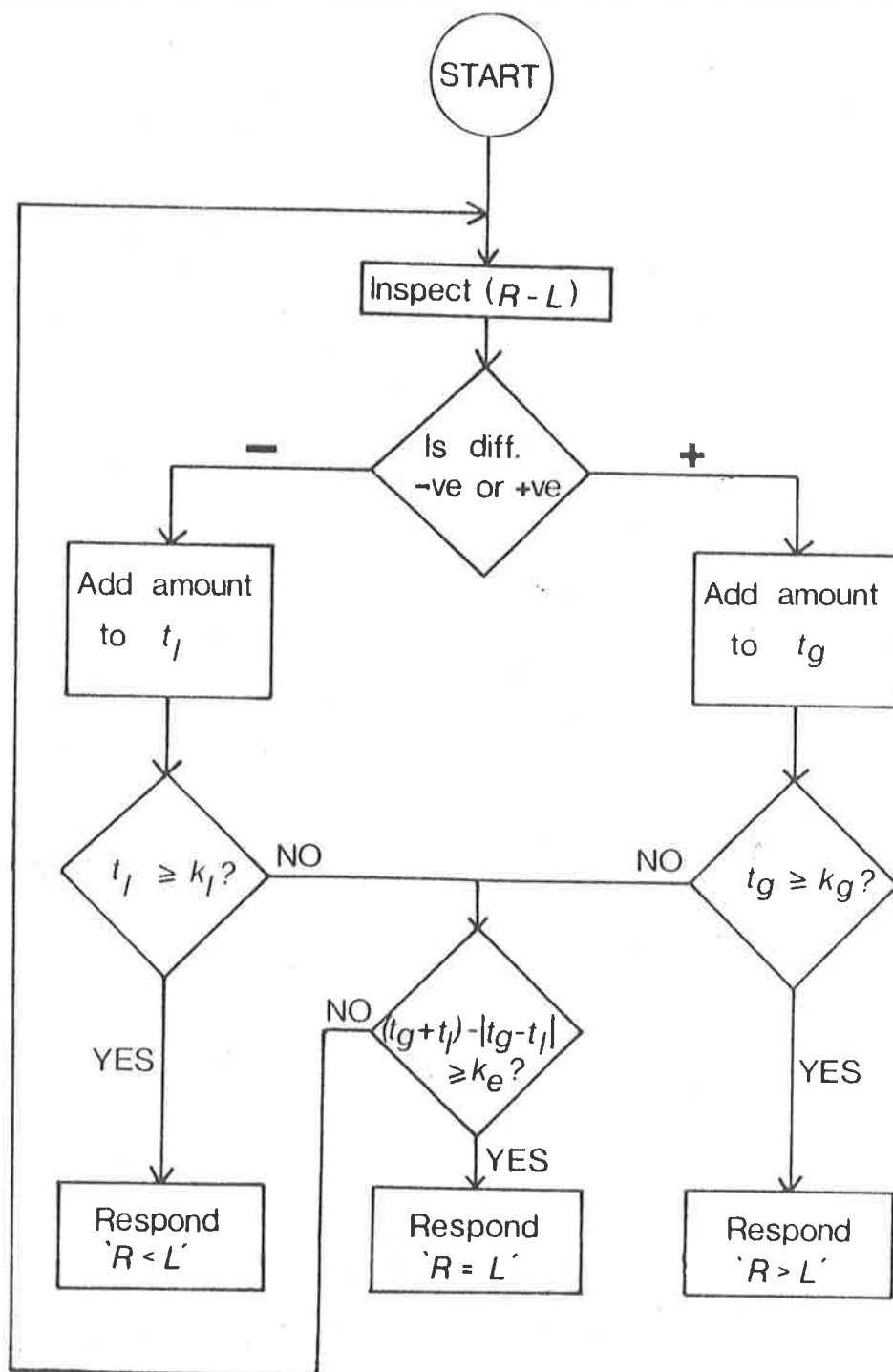


FIGURE 68 Flow diagram of the primary decision process in Vickers' adaptive accumulator model. The symbols are explained in the text. (Adapted from Vickers, 1978, figure 1.)

On the other hand, in a signal detection task, when the signal is typically an increment in intensity above the level of noise, the "greater" decision would lead to a signal response, while either the "equals" or (unlikely) "lesser" decision would trigger a nonsignal response. The "lesser" decision itself could be further suppressed by adopting a high value of k_1 .

Response times will be determined by the number of observations which must be taken before a critical total is reached. Since fewer observations will be required when stimuli are highly discriminable, that is, when $(R - L)$ differences are large, response latency will be inversely related to the level of discriminability. Also, since observations are assumed to occur at a fixed rate, at any particular level of discriminability response latencies for each response will be directly related to the critical total set for that response. The relative and absolute values of the critical totals will reflect bias and caution, respectively, in responding, which will thus affect the latency, as well as accuracy, of responses. At constant discriminability, setting a lower value on the critical total for a response, when bias increases in its favour, means fewer observations will be required to reach its critical total, so that the response will be made more quickly, as well as less accurately, since less evidence will now be required to trigger that response. If the critical totals for all responses are set at lower levels, when there is a decrease in the degree of caution in responding, then responses generally will be made more quickly, but at the expense of decreased accuracy. Conversely, more cautious behaviour, through adopting higher critical totals, will improve the accuracy of performance by requiring more evidence to be accumulated prior to any decision, but increase the time taken to make that decision.

When there is no bias in the critical totals, the confidence with which decisions are made is predicted by the 'balance of evidence' hypothesis (Vickers, 1972) in which it is proposed that confidence is based on the difference between the totals accumulated in favour of the alternative responses at the moment a decision is reached or a response is required. For instance, in a simple two-category task the confidence in a "greater" rather than "lesser" decision would be given by $(t_g - t_l)$. The hypothesis leads to a clear interpretation of the inverse relationship found between response latency and response confidence. When there is no relative bias, a response will be initiated after few observations only if the evidence has accumulated mainly in favour of one response alternative, and after many observations if evidence accumulates at a comparable rate for all response alternatives. In the former case, the short response time will correspond to a large discrepancy between the accumulated totals, and hence high confidence, while, in the latter case, the long response time corresponds to little difference between the accumulated totals, and hence to low confidence.

(b) The secondary control process

The confidence resulting from a decision is used to regulate the values of the critical totals in the primary decision process, the goal of this regulation being to make responses, on the average, with a certain target level of confidence. An independent process, identical in structure to that by which primary decisions are reached, is attached to each of the responses of the primary decision process. Each response is assumed to have a target level of confidence, CON_T , which the control process subtracts from the actual confidence, CON , with which each response is

made. Positive discrepancies, or 'overconfidence' in a response, can lead to a reduction in the value of k for a response, while negative discrepancies, or 'underconfidence', can lead to an increase in the critical total. These particular adjustments will restore confidence toward its target level, since the former adjustment tends to decrease the relative amount of evidence accumulated in favour of a decision whenever it is reached, and hence decrease its confidence, while the latter adjustment tends to increase the relative amount of favourable evidence that is accumulated.

The control process is illustrated in Figure 69 for the case of "greater" responses. Following each "greater" response, the target level of confidence for a "greater" response, $CON(G)_T$, is subtracted from the actual confidence of the response, $CON(G)$, and an amount, d , of under- or over-confidence results. As with stimulus differences in the primary decision process, amounts of overconfidence and underconfidence are accumulated separately, each positive difference being added to a total, T_{OG} (the total overconfidence in "greater" responses), and negative difference to a total, T_{UG} (the total underconfidence in "greater" responses). When a criterion amount of overconfidence, K_{OG} , or underconfidence, K_{UG} , has been accumulated, the value of k_g will be either decreased or increased, respectively, by an amount proportional to the difference between the accumulated totals. Similar accumulator processes are attached to the other response alternatives and it is assumed that whenever K is reached in any particular control process, so that k for the associated response is adjusted, only the values of T for that process will be re-set to zero.

In the control process a distinction can be drawn between bias and caution in terms of target levels of confidence in the same manner as with the critical totals in the primary decision process. Bias is manifested

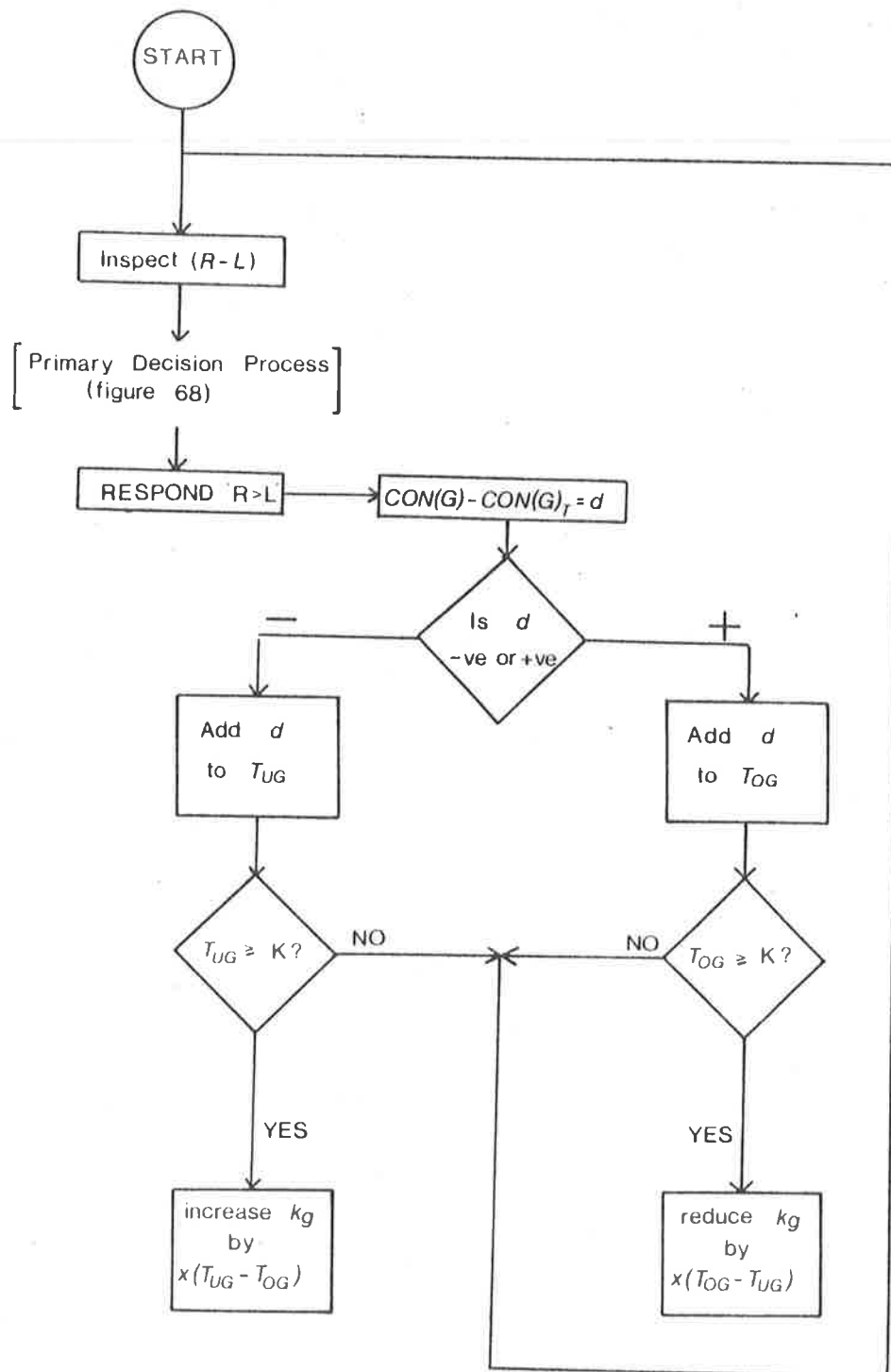


FIGURE 69 A branch of the secondary control process in Vickers' adaptive accumulator model. The critical total for a "greater" response, k_g , is reduced whenever the amount of overconfidence accumulated for the response, T_{OG} , exceeds a critical level, K_{OG} , and increased whenever the accumulated amount of underconfidence, T_{UG} , exceeds a critical level, K_{UG} . A coefficient, x , determines the proportionality of each adjustment to k_g . (Adapted from Vickers, 1978, figure 6).

as different values of target levels of confidence for the primary response alternatives, while caution is reflected as generally high or low values of the target levels. The effects on performance of inducing a bias toward a particular response, or varying the degree of caution exercised in responding, will be essentially the same, when mediated by target levels of confidence, as those expected to follow from change in the absolute and relative values of the primary decision criteria. For instance, if target levels of confidence are set at generally higher values, the discriminative capacity of the system is increased. This follows from increases in the critical totals for primary responses due to the upward adjustments necessary to maintain on average the higher target levels of confidence. As more evidence must be accumulated, in order to increase the absolute difference between critical totals when a decision is reached (and hence confidence), the accuracy of decisions will increase, again at the expense of increased response times. On the other hand, setting a low target level of confidence for a particular response (i.e. inducing a bias in its favour), will lead to a lowering of the critical total for that response, since decisions will initially be made with a predominance of overconfidence. The lowered critical total will thus decrease the amount of evidence necessary to elicit that response, thereby increasing its speed while decreasing its accuracy.

(c) Influence of stimulus discriminability and probability on decision criteria

In general, the level of discriminability at which stimuli are presented will influence the degree of caution the observer exercises in his responses, while unequal presentation probabilities will affect the bias shown in responding. This will be illustrated by use of a hypothetical two-category

decision task in which there are again two distributions of ($R - L$) sensory differences, one, r , associated with a positive mean difference and one, l , with a negative mean. The observer has to respond "R" if he decides that the sampled differences are associated with a stimulus in which R is objectively greater, and to respond "L" if he decides that they are associated with a stimulus in which L is greater.

(i) Stimulus discriminability. For a specific target level of confidence, response caution will be inversely related to the level of discriminability. This can be illustrated by referring back to Figure 59 in which one pair of stimuli, distributed as l_1 and r_1 , are of a greater magnitude than a second pair, distributed as l_2 and r_2 . Thus, if a series of trials in which observations are sampled from either l_1 or r_1 is followed by a series in which they are sampled from l_2 or r_2 , there will be a decline in the accuracy of L and R responses, since, on both l and r trials, there will be a greater likelihood of sampling evidence in favour of the inappropriate response. This will lead to a decline in response confidence for two reasons. Firstly, there will be an increase in the proportion of incorrect responses, and hence in those made at low levels of confidence. Secondly, even during those trials which lead to a correct response, evidence will tend to accumulate at more comparable rates in both accumulators, so that the balance in favour of an appropriate response will decline. There will then be a tendency for response confidence to fall below target levels, so that increases in both critical totals will be demanded. The resulting increase in response caution will not only restore response confidence toward its target levels, but, by increasing response accuracy due to more information being sampled prior to any response, it will also counteract to some degree the negative effect on performance of the decline in discriminability. For example, the

increase in the latency of left responses to the constant stimulus difference as discriminability declined during the previous experiment was consistent with an increase in the critical total for that response, although no other response measure gave an indication of increased caution.

(ii) Stimulus probability. If discriminability remains constant, but there occurs a decline in the relative probability of a stimulus, say the one associated with the x distribution, then, for a given number of trials, there will be a decrease in the number of correct R responses (those which occur with some fixed probability on x trials), and an increase in the number of incorrect R responses (those which occur with some fixed probability on l trials). This decline in the relative proportion of R responses made correctly means there will be a decrease in the proportion of R responses made when the sensory evidence is largely in their favour, that is, at higher levels of confidence, and an increase in those made when the evidence is not so favourable. Thus there will be a tendency for R responses to lead to underconfidence in the control accumulator process so that increases in the critical total for an R response become the more likely. Conversely, the increase in the proportion of L responses made correctly will lead to that response having a higher likelihood of overconfidence, so that its critical total will be lowered. Bias in favour of the more probable stimulus is thus predicted, which agrees with the direct relationship between response bias and stimulus probability specified, for example, in the 'ideal observer' hypothesis.

(iii) Bias and response confidence. An obvious deficiency in the 'balance of evidence' hypothesis might be apparent when bias exists, and this requires a more comprehensive formulation. When the critical total

for a response is lowered, the net effect should be to decrease the relative amount of possible evidence in its favour, thereby lowering confidence in that response. Consequently, confidence in a response would be expected to decrease as the bias toward it increases. However, the 'balance of evidence' hypothesis is strictly applicable only when no bias exists in the primary response criteria (Vickers, 1982). For a two-category task, confidence in response A , rather than B , is given in its general form by

$$CON(A) = (t_A - t_B) + (k_B - k_A) \quad (12)$$

where t_A and t_B are the accumulated totals in favour of the A and B responses, respectively, when a decision is reached, and k_A and k_B are the respective criterion totals. Confidence is thus seen as an assessment of posterior odds in terms of a Bayesian composite of a quantity, $(t_A - t_B)$, that measures the likelihood that a response was correct, given the available evidence, and a quantity, $(k_L - k_R)$, that is determined by the subjective prior odds that the stimulus associated with that response would occur, that is, the bias toward making that response. Thus, only when there is no bias, and $k_R = k_L$, does the 'bias factor' at the right of equation 12 reduce to zero and the 'balance of evidence' hypothesis apply. Otherwise the confidence in a response will be weighted, either positively or negatively, by a factor which indicates the bias toward making that response. This factor will then tend to increase the confidence of a response as the bias towards it increases.

2. Previous Results : Sensory Adaptation Versus Higher-Order Control

(a) Intervention by other control processes

The direct relationship between stimulus probability and response bias specified above agrees with the usual experimental findings, as well as providing an explanation for vigilance decrement when the observer encounters a drop in signal probability at the start of a task. However, it still runs contrary to the results of Experiments 1 to 4, in which bias was found to increase toward the response associated with the stimulus which declined in probability over time. However, Vickers (1979, Chapter 10) has considered the possibility that additional control systems, and reference variables other than response confidence, may exist, and that target levels of confidence may be regulated by these systems.

Specifically, it was suggested that the long-term probability of a response might act as a reference variable in a higher-order control system that will temporarily intervene whenever the recent probability of that response undergoes change. This proposal coincides with that presented in Chapter 4 (section A. 2. (b)), where the response stabilization mechanism was considered within such a hierarchy of control. Further, it is compatible with the general theory of human behaviour advanced by Powers, Clark and McFarland (1960a, 1960b), in which the output from a higher-order control system serves to set the reference levels in the control systems of the next lowest order. To extend this argument slightly, given the goals of response stabilization, it might represent a higher-order control system that will temporarily lower the target level of confidence sought in a response whenever there is a decline in the recent proportion of that response, or increase its target level of confidence when it becomes more frequent.

If the preceding argument is examined, in conjunction with the basis for response confidence that was presented in the previous section, then the patterns of performance associated with a response which increases in relative frequency, as well as an alternative response which decreases, will match the results obtained in Experiments 1 to 4. For a situation in which there is no initial bias, a reduction in the target level of confidence for a particular response, as was noted above, will lead to a reduction in its value of k , due to the increased relative probability of overconfidence in that response. This effect will be further enhanced by the additional response confidence which follows from a positive bias factor (equation 12). The lower critical total will lead to shorter response times, and an increased probability of triggering that response on any trial. Thus, the conditional probabilities of making that response will increase, as will its speed and confidence. These were the trends associated with reports of the stimulus which decreased in probability during the previous experiments, and since the converse trends can be expected if target level of confidence is raised for a response, the patterns of performance associated with reports of the stimulus which increased in probability are also duplicated.

(b) Change in point of sensory indifference

It would be far more parsimonious if the effects of a change in stimulus probability could be explained without invoking intervention by some other control process. This might be accomplished if one considers the relationship between stimulus probability and the mean sensory effect experienced by the observer over a series of trials. The potential importance of mean sensory intensity was demonstrated by Vickers and Leary (1982). These authors found the observed changes in β and response probabilities in Experiment 4 to fit very closely the patterns predicted by a

model in which, given a symmetric payoff matrix, and no prior information about changes in signal probability, the observer adopts as a decision cutoff the mean value of the sensory intensities which have been experienced, where this mean is cumulatively determined from the start of the experiment.

In many respects, this model bears a strong resemblance to Helson's (1947, 1948) concept of 'adaptation-level', that value of stimulus intensity which represents the combined effect of all intensities experienced within a given range. It corresponds to a point of sensory indifference, in that stimuli above this magnitude will elicit a greater, or positive judgment, and those below, a lesser, or negative judgment of relative magnitude. An example would be that value within a particular range of lifted weights at which the observer could make neither a "heavy" nor a "light" judgment. Adaptation level is assumed by Helson to vary according to an averaging process. For instance, the insertion of a large positive difference into a series of judgments of the size of the (positive or negative) difference between the magnitude of two stimuli, could result in a small positive difference evoking an "equals" judgment. Similarly, the same judgment could be evoked by a small negative difference following the insertion of a large negative difference into the series.

In Helson's view, adaptation-level is a point, above which sensory differences are judged to be positive, and below which, as negative. It is similar to the zero, or indifference, point along the hypothetical continuum of differences in sensory effect sampled in the accumulator process. This raises the possibility that, under conditions where signal probability or detectability vary, the sort of change that Vickers and Leary (1982) considered in regard to the observer's decision cutoff in a detection process might in fact occur to the value of this sensory indifference point for the

differences sampled in the accumulator process. That is, there occurs a sensory change, or adaptation, such that the indifference point will take the value of the mean sensory difference, cumulatively determined over the entire preceding trial sequence.

To return to the example of a two-category decision task, Figure 70 shows that, when the probability that L will exceed R on any trial increases from 0.5, that is, the probability of L increases with respect to r , the mean of all $(R - L)$ differences arising from l and r will take on a negative value. Alternatively, if the probability of L decreases from 0.5, then the mean $(R - L)$ difference will have a positive value. If the sensory indifference point is given by the mean difference experienced over all trials then, when probability changes, it will be located at a position like M^- or M^+ in Figure 70, rather than the zero-point at the intersection of l and r . The entire sensory

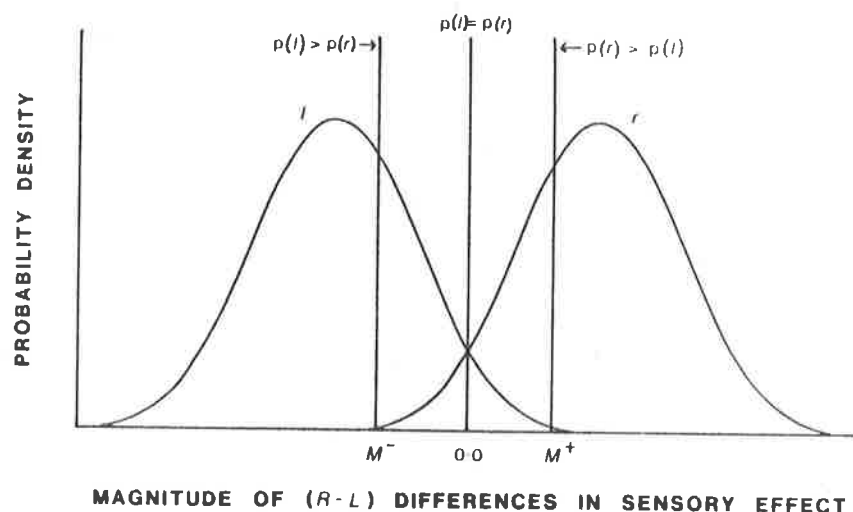


FIGURE 70 Hypothetical distributions of sensory differences produced by a stimulus with a greater magnitude to the right, distributed as r , and a stimulus with a greater magnitude to the left, distributed as l . The mean of all $(R - L)$ differences is 0.0 when the probabilities of l and r are equal, a negative value, M^- , when the probability of l exceeds r , and a positive value, M^+ , when the probability of r exceeds l .

continuum would need to be rescaled, but this could be achieved most simply by letting each new scale value equal its previous scale value minus a correction (equal to the value of the present indifference point along the previous scale). As a result, in the case of an increase in the probability of L , a boundary is placed at M^- in Figure 70, and the rescaled value of all differences below this boundary will be added to the negative total in the accumulator process, while the rescaled values of all differences above M^- will be added to the positive total. Since the rescaled differences are now more positive in value, the net effect is to decrease the relative amount of evidence likely to be sampled in favour of the L response on each trial, although the objective probability of L exceeding R has increased. Conversely, if the objective probability of R exceeding L increases, the evidence gathered in favour of R responses during each trial will tend to decrease. If such an effect does take place at the sensory level, and criterion levels are held constant, or adjusted only slowly, then the adaptive accumulator model could account for the patterns of performance found in Experiments 1 to 5 when signal probability and detectability varied.

(i) Signal probability. The initial effect of a decline in signal probability, as shown in section 1. (c). (ii), would be to decrease the proportion of correct signal responses and increase that of incorrect signal responses so that the criterion for a signal response will be raised and that for a non-signal response lowered. At the same time, however, the increased proportion of n trials will shift the sensory indifference point from the intersection of the s and n distribution toward some lower point along the sensory continuum, analogous to a shift from 0.0 toward M^- in Figure 70. Thus the evidence sampled in favour of a signal response on each trial will increase and that sampled in favour of a nonsignal response will decrease. The

resultant increased confidence of signal reports will make downward adjustments of the signal criterion become more likely, while the decreased confidence in nonsignal reports will make increases in the nonsignal criterion more likely. These trends will not only tend to diminish the initial bias against signal responses, but, since the indifference point varies according to the cumulatively-based, rather than local, signal probability, they would likely continue beyond the point at which the probability-bias effect has remedied the imbalance in confidence produced by the change in local response probabilities. If so, the indifference point will continue to move downward, increasing the evidence in favour of a signal response on each trial, so that the longer-term effect will be increases in the speed and confidence of signal decisions as well as the conditional probability of a signal response on any trial. Meanwhile, nonsignal responses will decrease in conditional probability, speed and confidence. Hence, on the basis of this model, whenever there was a reduction in the local probability of a signal during its ramp decline in Experiments 1 to 4, there would have ensued a relatively short period during which the bias toward a signal response tended to decrease. Superimposed on this, and continuing throughout the entire trial sequence, would be the dual effects of a tendency for more evidence to be sampled in favour of signal responses, that is, a reversal in the trends associated with decreased signal response bias, and a further reversal of the bias itself as the confidence in signal responses increased.

(ii) Signal detectability. A decline in the detectability of a signal would be equivalent, in the present context, to a decline in the mean ($R-L$) difference arising from the x distribution as the mean of the l distribution remains constant. This is shown in Figure 71 as a shift in x from x_1 to x_2 which, if the a priori probabilities of x and l are equal at 0.5, will shift

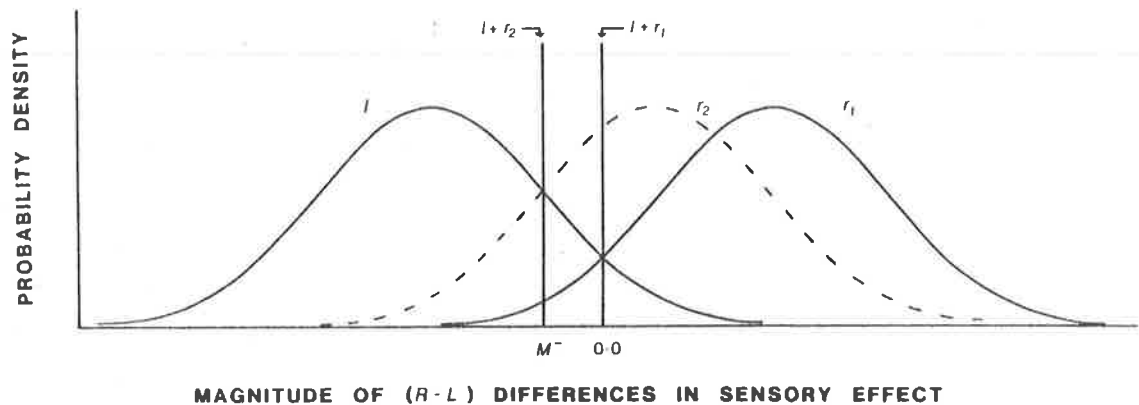


FIGURE 71 Hypothetical distributions of sensory differences produced by a stimulus with a greater magnitude to the left, distributed as l , and two stimuli with a greater magnitude to the right: one of the same absolute mean $(R-L)$ difference as l , distributed as r_1 , and one of a smaller mean difference, distributed as r_2 . When a priori probabilities are equal, the mean of all $(R-L)$ differences in l and r_1 combined is 0.0, and the mean of all differences in l and r_2 combined is some negative value, M^- .

the mean of all $(R - L)$ differences from 0.0 to some negative value, shown as M^- . In this case it should be noted that this negative value is again at the intersection of the two sensory distributions. Thus the indifference point will be shifted from that value associated with the intersection of the previous sensory distributions toward the value associated with the intersection of the distributions from which observations are presently being sampled.

If the sensory indifference point shifts from 0.0 to M^- , and performance during a sequence of trials in which r is distributed as r_2 is compared with a previous sequence in which r was distributed as r_1 , then change in performance will be the same as that associated with a symmetrical decline in the stimulus discriminability of both alternatives. That is, the proportion of both the l and r distributions which fall on the appropriate side of the indifference point declines so that the relative amount of

evidence sampled in favour of either inappropriate response will increase. This will lead to symmetrical decreases in response confidence and accuracy, and increases in response latency. These are the overall results found in Experiment 5 for both signal and nonsignal responses as signal detectability declined. There was, however, an initial period during which there was no change in performance during nonsignal presentations. This is consistent with the model, since the signal distribution for current trials would shift downward at a faster rate than the cumulatively-determined indifference point, whose movement along the nonsignal distribution would have had little initial effect on responses made during nonsignal trials.

(c) Sensory bias in a data accumulation model of signal detection

A point of sensory indifference at a position other than the intersection of the sensory distributions corresponds to a bias at the level of sensory processing. The possibility of such a bias within a data accumulation model of signal detection has already been considered by Pike (Pike, 1973; Pike & Ryder, 1972). Pike's 'multiple observations' model assumes that, during each trial, a separate count is made of the number of observations in favour of each response alternative until a critical count in favour of one of the alternatives has been reached, whereupon that response is initiated. The model allows for biasing effects at the sensory level as it shares the classical detection theory assumption of a critical observation magnitude, or criterion, along an observation axis of absolute sensory magnitude. Accordingly, in the proposed decision process, observations above this magnitude increment the count in favour of a signal response, and those below this magnitude increment the count for a nonsignal response. Thus, for each discrete observation during a trial, the probability that

the counter associated with the correct response will be incremented is given by the area of the relevant sensory distribution that lies on the appropriate side of the decision criterion, and the area lying on the opposite side gives the probability that the incorrect counter will be incremented. If there is no sensory bias, the criterion is placed at the intersection of the two sensory distributions and the relative probability that any observation will increment the appropriate counter is the same for both distributions. When there is sensory bias, corresponding to what has been proposed as a shift in the sensory indifference point from its intersection position, the probability that any observation provides evidence in favour of the corresponding correct response will depend upon which sensory distribution is being sampled.

Pike (1973) sees response bias in terms of the relative values of the critical counts for the two response alternatives and assumes that "the observer keeps a precise and constant control over the critical count for any given experimental condition" (p. 61). However, no mechanism is specified by which bias is introduced into the decision process or by which the observer maintains control over the values of the critical counts. Further, while the manner in which sensory bias may come to exist is also not specified, and decision bias is considered to be mainly embodied in the counting process, Pike and Ryder (1973) proposed that a change in the underlying sensory criterion was responsible for previous findings that signal response latencies increased with decreasing signal probability. Specifically, the probability that an observation would increment the signal counter was assumed to be high at high signal probability, whereas the probability of a signal count was low at the low signal probability, the respective consequences of a low and high criterion in a classical signal

detection model. Thus, it seems safe to conclude that the sensory criterion is assumed to change in the same direction as the optimal decision criterion in signal detection theory. Such changes would, of course, produce the opposite trends to those found when stimulus probabilities varied during the trial sequences of the previous experiments. This follows since a decline in signal probability accompanied by an increasing sensory criterion would continually decrease the probability that the signal counter would be incremented on both signal and nonsignal trials, thereby decreasing the conditional probabilities of signal reports rather than, as was found, increasing them.

The present model, on the other hand, sees sensory bias as arising from an adaptation process which is, in all respects, beyond the control of the observer. The sensory indifference point, except when a priori probabilities are equal, will be located at a position on the opposite side of the intersection of the sensory distributions than the optimal position of the decision criterion specified by signal detection theory. Yet, when a priori probabilities differ, but remain constant, the direct relationship between response bias and stimulus probability can still be explained since training and instructions would normally lead the observer to seek a lower level of confidence in reports of the stimulus known to be more probable. Thus, while less evidence accrues in favour of the more probable stimulus, this is offset by the lesser amount required to reach the target level of confidence sought in reports of that stimulus.

3. A Dual-Criterion Self-Regulating Signal Detection Model

The 'multiple observations' detection model described above has an advantage over traditional signal detection models in that response bias and caution are both featured in the decision process. However, since they are embodied as the relative and absolute values, respectively, of the critical counts, the model bears more similarity to the general class of optional-stopping data accumulation models than to the traditional detection model. Nevertheless, a detection model can be developed in which response caution, as well as bias, is specified in terms of decision criteria placed along a continuum of sensory activity, or monotonic transformation of that activity. Such a model is further relevant to the present discussion, since both bias and caution are regulated by a feedback process. Also, while multiple observations can occur during a trial, a response is based upon the evidence contained in a single observation, as is supposed by the traditional detection model.

(a) The decision process

The underlying decision process is given by the 'interval of uncertainty' model of Pike (Pike, McFarland & Dalglish, 1974; Pike & Dalglish, 1976), an extension of the sequential decision model of Swets and Green (1964). According to the model, the observer makes repeated observations, x , of sensory input until the evidence contained in a single observation has exceeded the criterion level for a response. Hence, there is a critical observation magnitude for both a signal and nonsignal response so that, as shown in Figure 72, two decision criteria are located along the observation axis: one, x_s , above which the observation is classified as having arisen from a signal, and one, x_N , below which the observation

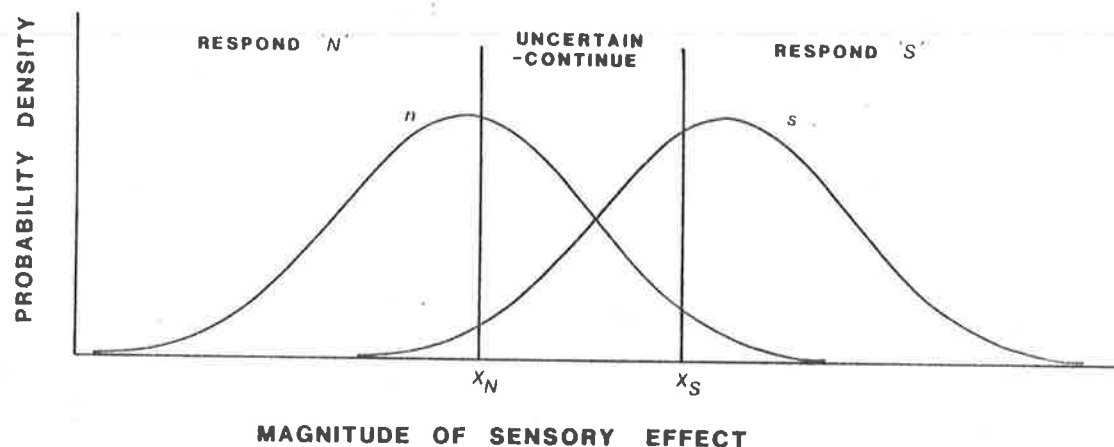


FIGURE 72 Hypothetical distributions of the sensory effect produced by signal, s , and nonsignal, n , stimuli in the 'interval of uncertainty' model. When the magnitude, x , of an observation falls below x_N , the observer makes a nonsignal response (N), and when it exceeds x_S , the observer makes a signal response (S). When x falls between these critical values the observer cannot reach a decision and so makes a further observation.

is classified a "nonsignal". The range of values of x bounded by these criteria, the interval of uncertainty (IU), does not provide the observer with enough evidence to reach a decision so that, when one of these values is first sampled, a further observation will be required. Observations then continue until one occurs outside the IU, whereupon the associated response is triggered.

In Pike's model the sequence of observations is assumed to occur during a single presentation of a stimulus, whereas the original sequential model was concerned with decision behaviour during a series of discrete presentations of the same stimulus. It should also be noted that the IU model is a special version of the original model, since it further assumes that the decision to respond is based upon the information contained in a single observation. That is, information is not integrated over the series of observations. This latter assumption appears satisfactory for the

application of the model, at least in situations where the observer has not been encouraged to integrate evidence over a series of discrete stimulus presentations (Swets & Green, 1964), or at some time been required to give a likelihood rating at each stage of such a series (Swets & Birdsall, 1967).

(b) Response probability, latency and confidence

Since an observation can lead to a correct, incorrect, or uncertain ("continue") decision, there are three decisional probabilities conditional upon each sensory distribution. These conditional probabilities will determine the empirical probabilities of a detection, missed signal, correct rejection and false alarm, as well as the expected number of observations which occur prior to each response and, thus, response latency (the proofs are given by Pike et al., 1974). For example, in Figure 72 the area of s below x_N gives the conditional probability that an observation drawn from s will lead to an incorrect nonsignal decision, $p(N/s)$; the area of s between x_N and x_S gives the probability of an uncertain decision, $p(U/s)$, so that another observation is required; and the area of s beyond x_S gives the probability that an observation of s will lead to a correct signal decision, $p(S/s)$.

The observed probability of a detection is then given by

$$p(\text{detection}) = \frac{p(S/s)}{1 - p(U/s)}, \quad (13)$$

and a missed signal by

$$p(\text{miss}) = \frac{p(N/s)}{1 - p(U/s)} \quad (14)$$

Similarly, the area of n below x_N , $p(N/n)$, between x_N and x_S , $p(U/u)$, and above x_S , $p(S/n)$, will determine the probability of a correct rejection, given by

$$p(\text{correct rejection}) = \frac{p(N/n)}{1 - p(U/n)}, \quad (15)$$

the probability of a false alarm, given by

$$p(\text{false alarm}) = \frac{p(S/n)}{1 - p(U/n)}. \quad (16)$$

The expected number of observations prior to either a detection or missed signal, that is, the mean latency of responses on signal trials (in observation units), is given by

$$\text{latency}(s \text{ trials}) = \frac{1}{1 - p(U/s)}, \quad (17)$$

and the mean latency of correct rejections and false alarms, in observation units, as

$$\text{latency}(n \text{ trials}) = \frac{1}{1 - p(U/n)}. \quad (18)$$

Provided that differences in latency arise from differences in observing time only, equations 17 and 18 predict that the mean latencies for correct and incorrect responses to a stimulus will be identical, a prediction that is not supported empirically (e.g. Carterette, Friedman and Cosmides, 1965; Emmerich, Gray, Watson & Tanis, 1972). However, if it is assumed that some different amount of time is required in order to classify a stimulus once an observation occurs outside the IU than is required within the IU by the decision to continue, and that this

classification process is shorter the greater the amount of evidence an observation contains in favour of the successful response, then a distance-from-criterion function of response latency can be specified in the same manner as has been done previously in the single-criterion detection model. This is shown in Figure 73 as decreasing latency as the distance from a particular decision criterion is increased. Further, the distance an observation lies outside the IU might also be deemed to indicate the certainty with which a response is initiated, as that distance represents the degree to which the evidence contained in an observation exceeds the amount associated with no response certainty. Hence, response confidence will again be considered a direct function of the distance from the decision criterion of the observation which triggers a response, as

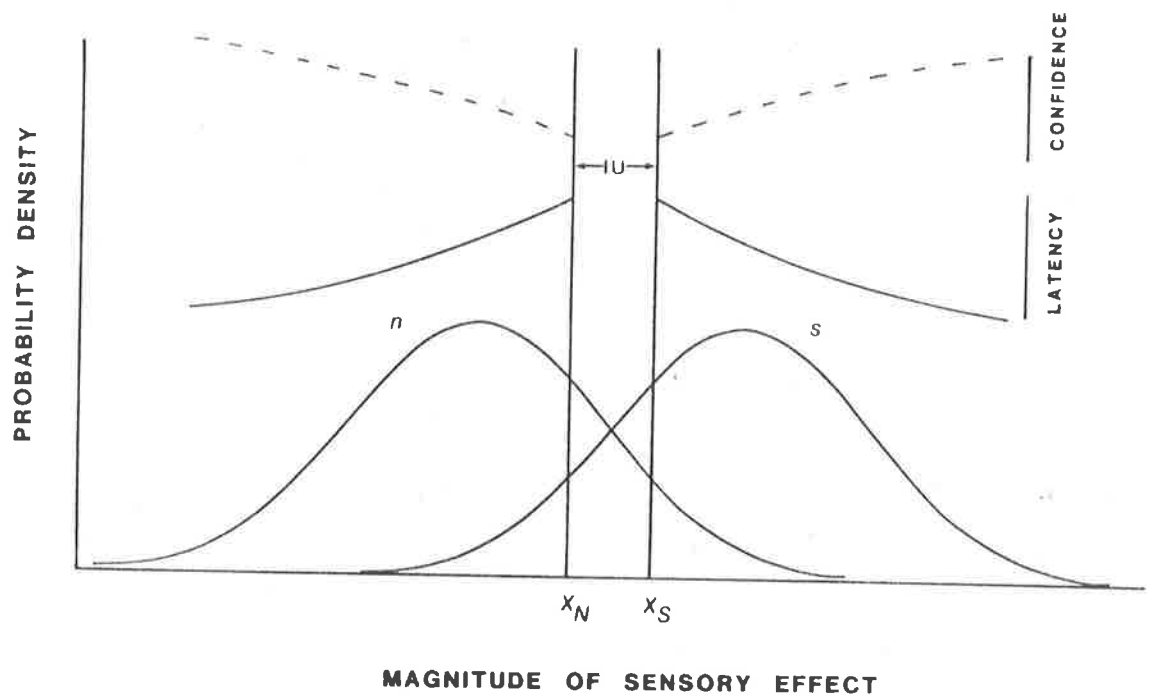


FIGURE 73 The distance-from-criterion hypothesis applied to the 'interval of uncertainty' (IU) detection model. The unscaled latency required to classify an observation, x , as a signal (S) or nonsignal (N) is given by the unbroken line, and the unscaled confidence resulting from that decision is given by the dashed line.

also shown in Figure 73. Thus, in line with the usual finding, the more probable (usually the correct) response to a stimulus will be faster and more confident than the less probable (usually the incorrect) response.

The response probability-latency relationship could also be produced if it is assumed that more than one favourable observation outside the IU is required for a response, as, for example, in Pike's (1973) multiple observations model. Thus, the more probable response would usually require fewer observations due to the greater likelihood of observations being sampled in its favour. Further, if an accumulator process such as Vickers' (1979) is assumed, in which the evidence contained in each observation is stored, then, whenever a response is initiated, the evidence in its favour, and thereby confidence, will usually be greater for the more probable response to a stimulus. This follows, since the average amount of evidence contained in a favourable observation, in terms of the extent to which it exceeds the criterion, is greater for the more probable response. However, these possibilities will not be considered further, since they add to the basic detection model the characteristics of an optional-stopping accumulator process - and a very unwieldy one - as both bias and caution could be manifested at the sensory as well as decision levels. To provide the greatest contrast between traditional detection models and data accumulation models, only the basic dual-criterion detection model will be considered, with, for whatever underlying reason, the additional assumptions of a distance-from-criterion function for both response latency and confidence.

(c) Regulation of response bias and caution

The location and width of the interval of uncertainty reflect bias and caution, respectively, in responding. When there is no bias toward either response, the IU will be located at a symmetrical position with respect to the s and n distributions, as in Figures 72 and 73. When there is signal response bias the IU is placed further to the left along the sensory continuum, as shown in Figure 74, so that comparatively less evidence is required for a signal response and more for a nonsignal response. In the same way a nonsignal response bias will place the IU to the right of its symmetrical location. A general increase in caution will lead to the expansion of the IU in both directions, that is, the amount of evidence required prior to any response will increase, thereby displacing the signal criterion to the right and nonsignal criterion to the left. As an illustration, Figure 72 indicates a higher degree of caution than Figure 73.

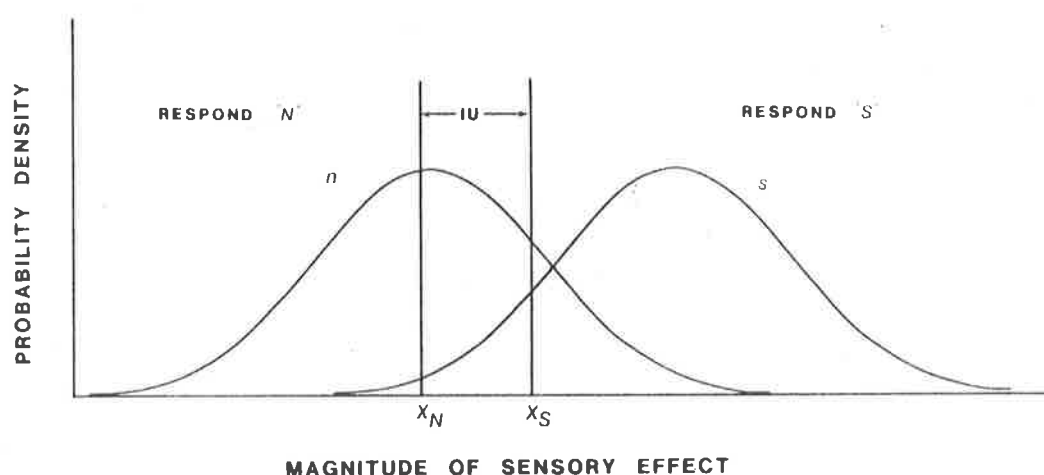


FIGURE 74 Hypothetical distributions of signal, s , and nonsignal, n , stimuli, and the positions of the criteria for signal, x_S , and nonsignal, x_N , responses when there is bias toward making signal responses.

The predicted effects of change in bias or caution are largely consistent with those of the accumulator model, provided the distance-from-criterion relationships are assumed. When there is bias toward reporting a particular stimulus, the proportion of the distribution associated with that stimulus which is contained within the IU will be less than the proportion of the distribution associated with the alternate stimulus. Thus, in the case of a signal response bias, as can be seen in Figure 74, the probability of an uncertain response, given an observation of s , is less than that of an uncertain response, given n , while the probability of a correct response is higher given s . By equations 13 to 18, responses during s trials will then be more accurate and faster than those on n trials. Also, since the mean observation distance of signal responses is greater than that of nonsignal responses, for both correct and incorrect responses, signal responses will be made more confidently than nonsignal responses. The opposite effects would hold for the case of a nonsignal response bias.

An increase in caution will again lead to more accurate, but slower, responses. As a comparison of Figures 72 and 73 will indicate, an increase in the size of the IU will increase the proportion of both distributions which lie inside it, leading to larger expected number of observations prior to a response, but a higher likelihood of a correct response due to this increased sampling. For example, Table 31 gives the expected proportions of correct and incorrect responses, number of observations prior to a response, and signal detection measures, when there is an increase in an IU placed at a symmetrical location with regard to a s and n distribution whose means are separated by two standard deviations. As shown in the Table, as the interval expands from a width of 0.0 (the single-criterion case), an increase in the mean number of observations taken prior to a response is accompanied by an

TABLE 31. Expected performance measures when the means of the s and n distributions are separated by 2.0 standard deviation units and an unbiased IU is expanded from a width of 0.0 (single criterion) to 3.0 standard deviation units.

Width of IU (s.d. units)	Response Probability		Mean Latency (observation units)	Signal Detection Measures	
	Correct	Incorrect		β	d'
0.0	.84	.16	1.0	1.0	2.0
0.5	.88	.12	1.14	1.0	2.34
1.0	.91	.09	1.32	1.0	2.71
1.5	.94	.06	1.56	1.0	3.06
2.0	.96	.04	1.91	1.0	3.42
2.5	.97	.03	2.42	1.0	3.79
3.0	.98	.02	3.17	1.0	4.16

increase in the probability of a correct response, so that measured sensitivity, d' , increases from 2.0 (its single-criterion value).

A discrepancy emerges between the predictions of the accumulator and dual-criterion detection model when the relationship between degree of caution and response confidence is considered. As the IU expands symmetrically, the average distance at which observations lie on either side of it will decrease and, hence, so will the confidence with which responses are made. This inverse relationship between caution and response confidence has important implications for the model if, as was suggested for the single-criterion model, the observer utilizes the distance that observations lie from the decision criterion in order to control the position of that criterion. In the case of the dual-criterion model, the goal of such regulation would be to make observations at some constant average distance from the interval of uncertainty. This model offers considerably more flexibility than the single-criterion model: the IU can be contracted or expanded and, at the same time, shifted in either direction along the observation axis. Thus, it can counteract symmetrical as well as asymmetrical change in the mean observation distances of the two responses.

It will be assumed that the criteria positions are determined by initial task experience and instructions, and that the mean observation distance which results on each side of the IU is used as a reference variable in the subsequent regulation of the associated decision criterion. Thus, if equal importance is attached to two equiprobable responses, the IU will be placed at a symmetrical location and the mean observation distance which is sought for a response will be the same on either side of the IU. When less evidence is required for one response than for its alternative, whether this bias is produced by unequal a priori probabilities or the type of payoff matrix, the criteria will be placed at higher or lower positions along the observation axis, and the mean observation distance maintained beyond one criterion (high bias) will be greater than that maintained beyond the other criterion. Any manipulation which subsequently causes mean observation distances to vary will produce adjustments of the criteria toward positions which will restore the mean observation distances to their previous (target) values. A decline in signal probability, for instance, will cause signal responses to be triggered by observations which lie at a closer average distance to the signal criterion, and nonsignal responses by observations which lie further from the nonsignal criterion, compared to previous observation distances. An increase in bias toward making signal responses is then demanded, which is achieved in the present model by adjusting both the signal and nonsignal criterion downward.

It is when symmetrical effects are produced on the mean observation distances of the two responses, so that caution must be adjusted, that the predictions of the model differs most radically from those of the adaptive accumulator model. This is the consequence of the inverse relationship between degree of caution and mean observation difference.

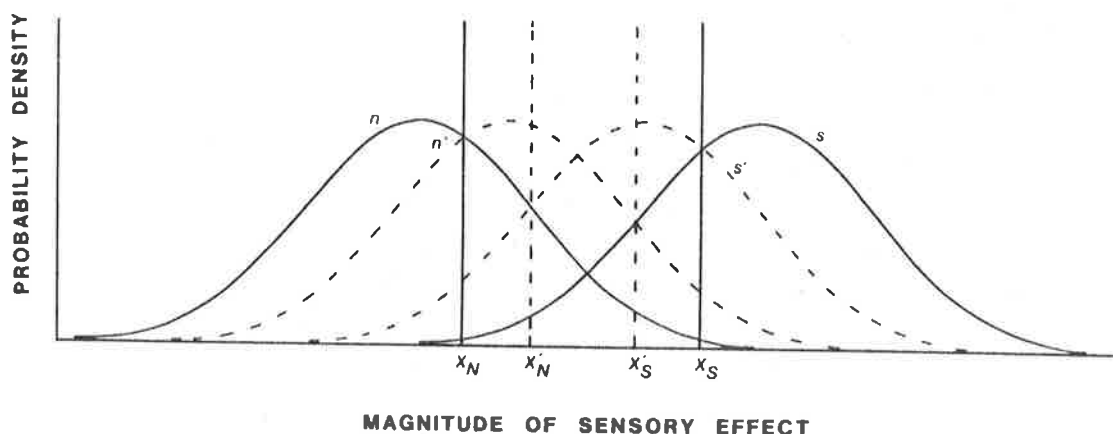


FIGURE 75 Hypothetical distributions of s and n when there is a decline in the magnitude of s , so that it becomes distributed as s' , and an increase in the magnitude of n , so that it becomes distributed as n' . The two criteria, x'_S and x'_N , will produce the same mean observation distance beyond the interval of uncertainty when s' and n' are being sampled, as is produced by x_S and x_N when s and n are being sampled.

In Figure 75 are shown the hypothetical distributions of s and n when a decrease in the magnitude of signal stimuli is accompanied by an equivalent increase in that of nonsignal stimuli. Since the mean observation distance will decrease to the same extent for both signal and nonsignal responses, the nonsignal criterion must be adjusted upward, and signal criterion downward, in order to decrease the mean observation distances and make responses at their previous levels of confidence. That is, caution must be decreased. The shifts in the two distributions are equivalent to those which occurred in the previous experiment when the general level of discriminability declined, so that the adaptive accumulator model would predict an increase, rather than decrease, in caution. Thus the models are directly opposed in their predictions about the effects of a fall in the level of discriminability. The accumulator model would demand an

increase in the value of the primary decision criteria, and hence caution, in order to increase the maximum possible balance of evidence in favour of a response, and so restore confidence toward its target levels. The dual-criterion detection model would demand a contraction of the IU, a decrease in caution, in order to increase the distance at which observations lie outside the IU, and so achieve the same goal of an increase in response confidence. Due to the symmetrical nature of the changes which occur when discriminability is varied, there will be no confounding effects on the predicted response of either model due to shifts in the point of sensory indifference, or, perhaps, decision criteria, in the same direction as change in mean sensory effect.

B. EXPERIMENT SEVEN : CRITERION ADJUSTMENTS FOLLOWING
A STEP CHANGE IN STIMULUS DISCRIMINABILITY

The previous experiment failed to provide enough evidence of a change in caution to confirm the predictions of either model. However, the gradual nature in which discriminability was varied in that task may have made the measurement of such adaptive behaviour highly unlikely. Further, a very simple discrimination judgment had been required of the observer and resulted in high levels of response accuracy at all discrimination levels. Thus, the present experiment employed abrupt step changes in the general level of discriminability, rather than the ramp change of the previous task. Also, the discrimination judgment required of the observer was made more complex and more closely resembled the detection task used for Experiments 4 and 5. A pair of stimuli of constant magnitude was again presented throughout the task in order to determine whether

there would be any adjustment of the observer's decision criteria to compensate for the effects of variation in discriminability. If there is no adjustment, then performance at this constant stimulus magnitude will be the same when the stimulus is presented in a series with stimuli of greater magnitude as when it is presented in a subsequent series with stimuli of lesser magnitude. If there is adjustment, then, according to the adaptive accumulator model, responses to this stimulus will be more accurate and confident, but made more slowly, when there is a drop in the general level of discriminability. In contrast, the dual-criterion detection model predicts less accurate, but faster and more confident, responses.

1. Method

(a) Task and stimuli

On each trial two 96 by 116 mm displays, separated by a horizontal distance of 15 mm, were presented on the screen of a television monitor and viewed at a distance of approximately 4.0 metres. Each display comprised a 5 by 5 matrix of horizontal lines which varied in length, and the task of the observer was to decide whether the average line length was greater in the display on the right or the one on the left (see Figure 76). The lines in one of the displays, the 'standard', had a mean length on each trial of 10.27 mm, and range of 4.4 to 16.13 mm. The lines in the other display, the 'variable', assumed one of five possible longer mean lengths and associated ranges, and thereby produced one of five distinct (variable-standard) mean "line length differences". At the lowest line length difference of 1, the variable had a mean line length of 11.0 mm and a range of 5.87 to 16.13 mm; at 2 the lines had a mean length of 11.73 mm

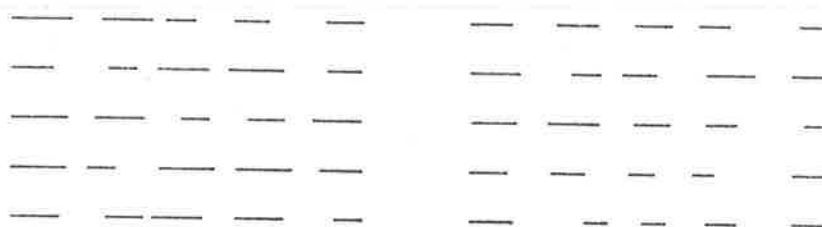


FIGURE 76 An example of a variable (left) and standard (right) display as was presented to the observer on each trial of Experiment 7. In this example the "left" response would have been correct. The above displays are not drawn to scale (see text for dimensions).

and range of 5.87 to 17.6 mm; at 3 the lines had a mean length of 12.47 mm and range of 7.33 to 17.6 mm; at 4 the lines had a mean length of 13.2 mm and range of 7.33 to 19.07 mm; and at the largest line length difference of 5, the mean was 13.93 mm and range from 8.8 to 19.07 mm. These differences in the mean line length of the two displays correspond to visual angles from $.011^\circ$, at line length difference 1, to $.053^\circ$ at line length difference 5.

At any time during the task, the observer was presented with only those line length differences contained within either a 'low', 'high', or 'median' level of discriminability. The low level of discriminability comprised the three smallest line length differences, 1, 2 and 3. The high discriminability level comprised the three largest line length differences, 3, 4 and 5. The median discriminability level comprised line length differences, 1, 3 and 5, so that, on the average, the level of discriminability here was intermediate between the low and high discriminability levels. The line length difference of 3 was presented at all three discriminability levels, which allowed the measurement of performance at a constant stimulus difference.

(b) Apparatus

Stimuli were generated by means of a graphics program contained in a D.E.C. PDP 8/F computer. The side on which the variable display was presented, and the line length difference on each trial, was determined by a disk-based D.E.C. PDP8/L computer. The display was visible for 2 seconds, unless terminated by a decision response. Following stimulus termination, there was a 2 second interval prior to the next stimulus presentation, during which a confidence response could be recorded, as well as a primary decision response if one had not terminated the previous stimulus. The responses were classified by the second computer, which also measured the latency of the primary decision response and value of the confidence response, and provided a summary of these statistics after each block of training trials and at the end of the main experimental session.

Observers used the index finger of their dominant hand to depress the left or right of two perspex buttons, lying on either side of a wedge-shaped resting position, to indicate whether the display on the left or the one on the right, respectively, was judged to have the longer average line length. Confidence was then expressed along a five-category scale by depressing one of a line of ten horizontal buttons, divided in the middle by a yellow line so that five were in front of each primary response button. For both primary responses the confidence scale ran from "least confident" at the innermost button, lying next to the yellow line, to "most confident" at the outermost button.

(c) Design

An experimental session consisted of two blocks of 48 training trials each, followed by 672 consecutive test trials. There was a gap of about three minutes after each block of training trials, during which a print-out

of performance was examined. The main experimental session, which could last for a maximum of about 45 minutes, was divided into fourteen blocks of 48 trials each, presented as an unbroken sequence of trials. During each training and test block, stimuli were presented at a single level of discriminability, with line length differences occurring in random order with the constraint that an equal number of stimuli, sixteen, be presented at each of the three line length differences for that particular discriminability level. The variable display appeared to the left and right in random order, with the constraint that it appear an equal number of times on each side at each line length difference within each block.

The median discriminability level was employed during training in order to give observers experience of the range of line length differences to be presented. The main experimental session was divided into seven pairs of two trial blocks each. The discriminability level employed in each pair of trial blocks is shown in Figure 77. As can be seen in that figure, observers began the task at the same intermediate level of discriminability

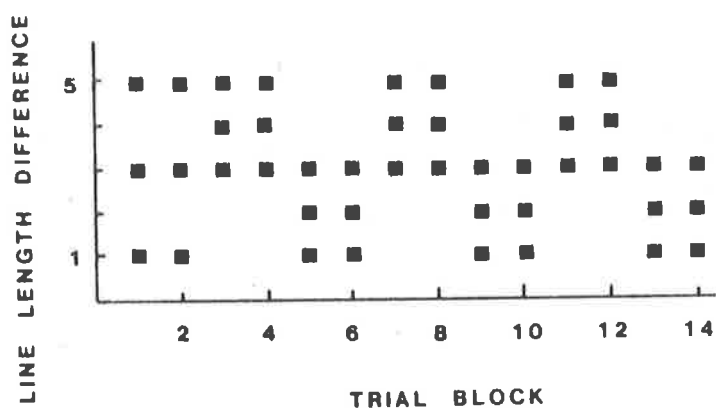


FIGURE 77 The line length differences which were presented during each trial block of Experiment 7. Observers performed under the median level of discriminability (differences 1, 3 and 5) over the first two trial blocks, after which discriminability alternated between its high (differences 3, 4 and 5) and low level (differences 1, 2 and 3).

that was encountered in practice, after which discriminability alternated between its high and low levels.

(d) Procedure

Observers were tested individually and in complete isolation. Watches were removed prior to the first block of training trials. Instructions were given which explained the nature of the discrimination task, and an example of one of the stimuli was shown to the observer. The use of the response buttons and confidence scale was explained, as was the need to make both responses within four seconds of stimulus onset. Observers were told that the discriminability of the task would vary during the main session, but not the exact nature of this variation, and also were told that during the practice trials stimuli would be randomly presented from "a high, a medium, and a low level of difficulty". Prior to the practice trials, observers were instructed to attempt to spread their confidence ratings over all five buttons for each primary decision response. Prior to the main session, additional instructions were given, which stated that, if it appeared at any time that the majority of confidence responses were being made at either of the extremes of the confidence scale, not to worry, but make sure that each confidence response was based on the immediate past decision, and not on any more general feeling of confidence.

(e) Observers

Twenty-eight students (14 male, 14 female), enrolled in first year Psychology at the University of Adelaide, volunteered to participate in the experiment as part of their course requirements. Ages ranged from 17 to 29 years with a median age of 18 years. All observers had normal or corrected vision.

2. Results

As in the previous experiment, analyses were aimed at determining, firstly, whether there was an effect of discriminability level on performance measures averaged across all line length differences at each level of discriminability and, secondly, whether there was an effect of discriminability level on performance at the line length difference of 3 which was presented throughout the task. The former analysis sought to determine whether performance had been sensitive to the step changes in discriminability so that any findings of the second analysis could be attributed to this variation. Data from the first two trial blocks, in which stimuli were presented at the median discriminability level, were excluded from the analysis, as this phase of the task was intended to stabilize performance prior to the shifts in discriminability. Thus the factor of discriminability level took on only the high and low value. A second factor, period, was defined in order to roughly determine the change over time in performance following a shift in discriminability. Data were classified here according to whether the measurement had been taken in the first or second block of 48 trials following each step change in discriminability.

(a) Performance across all line length differences

In this section is reported the results of analyses performed on data which had been averaged across the three line length differences which had been presented in each block of trials.

(i) Response probability. The trends which occurred during the session in the probability of making a correct report were unrelated to whether the stimulus had longer lines to the right or to the left, and are therefore shown in Figure 78 for the probability of making a correct response

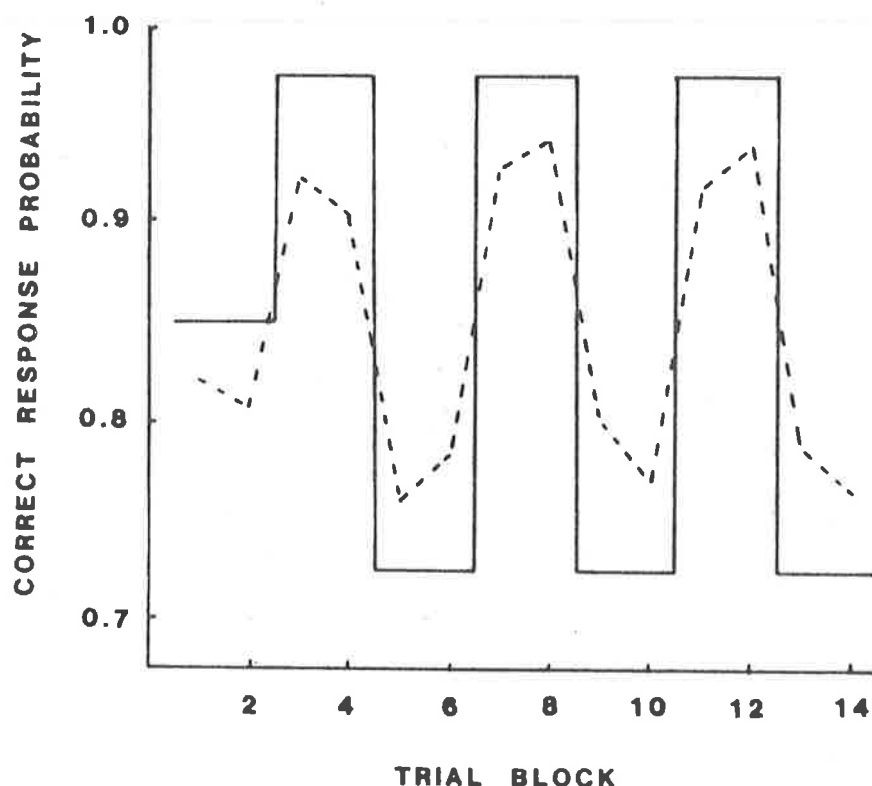


FIGURE 78 The mean probability of making a correct response across the line length differences in each block of 48 trials. The unscaled level of discriminability is given by the unbroken line.

with left and right responses combined. The figure suggests that performance accuracy was strongly affected by the step changes in discriminability and, as would be expected by both theoretical models, an increase in discriminability consistently produced an increase in the proportion of correct responses while a decrease in discriminability always led to a reduction in the proportion of correct responses. However, over the first two shifts in discriminability the changes in accuracy over the 96 trials which followed a step increase and decrease in discriminability were different from those which followed a step increase and decrease over the final four shifts in discriminability (Figure 78).

Specifically, the first step increase in discriminability was followed by a drop in the probability of a correct response from its initially higher value, and the first step decrease by an increase in the probability of a correct response from its initially lower value. These trends correspond to the decreases in accuracy at high discriminability, and increases at low, that the accumulator model predicts will follow from a decrease and increase, respectively, in caution in response to a step change in discriminability. However, over the final four shifts in discriminability these trends were reversed: correct response probability continued to increase during sequences of high discriminability, and continued to decrease at low discriminability. That is, these later trends reflected the respective effects of increased caution at high discriminability, and decreased caution at low discriminability, predicted by the dual criterion detection model. As a result of this dichotomy, analyses were performed on the data from the final four shifts in discriminability separately from that for the initial two shifts.

The results of the analyses of variance on correct response probability are given in Table 32. The main effect of discriminability was significant over the first two shifts in discriminability, trial blocks 3 to 6, as well as over the final four shifts, trial blocks 7 to 14. There was no main effect of period in either analysis, due in each case to the opposing trends which followed a step increase and decrease in discriminability, and which led to the significant interaction effects. Subsequent analyses of the simple effects of period at each level of discriminability revealed that, over the first two shifts in discriminability, neither the decrease in correct response probability at high discriminability, or its decrease at low discriminability, was quite strong enough to reach significance ($F = 3.072$, d.f. = 1/27, $p = .09$, and $F = 3.104$, d.f. = 1/27, $p = .09$, respectively). Over the final four shifts,

TABLE 32. Results of analyses of variance on correct response probability averaged across the three line length differences in each block of trials.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Over trial blocks 3 to 6</u>				
	Discriminability (D)	1	1.673	174.512	.00001
	D x S	27	.009		
	Period (P)	1	.000	.846	
	P x S	27	.007		
	D x P	1	.041	6.588	.05
	D x P x S	27	.006		
(b)	<u>Over trial blocks 7 to 14</u>				
	Discriminability (D)	1	3.747	298.797	.00001
	D x S	27	.012		
	Period (P)	1	.001	.349	
	P x S	27	.005		
	D x P	1	.080	9.564	.005
	D x P x S	27	.008		

however, both the increase at high discriminability, and decrease at low discriminability, were significant ($F = 9.749$, $d.f. = 1/27$, $p < .005$, and $F = 5.074$, $d.f. = 1/29$, $p < .05$). Thus, the two interactions indicate that there were definitely opposing trends following a step increase and a step decrease in discriminability, and that these trends reversed after the first two shifts in discriminability. The simple effects of period indicate that the changes in correct response probability over the first two shifts, which imply the *inverse* relationship between caution and discriminability specified by the accumulator model, were not as strong as the changes which occurred over the final four shifts in discriminability, and which imply the direct relationship specified by the dual-criterion model.

(ii) Response latency. The same trends were indicated in the latencies of correct and incorrect responses to both types of stimulus, as shown in Figure 79, where latencies have been averaged across all responses

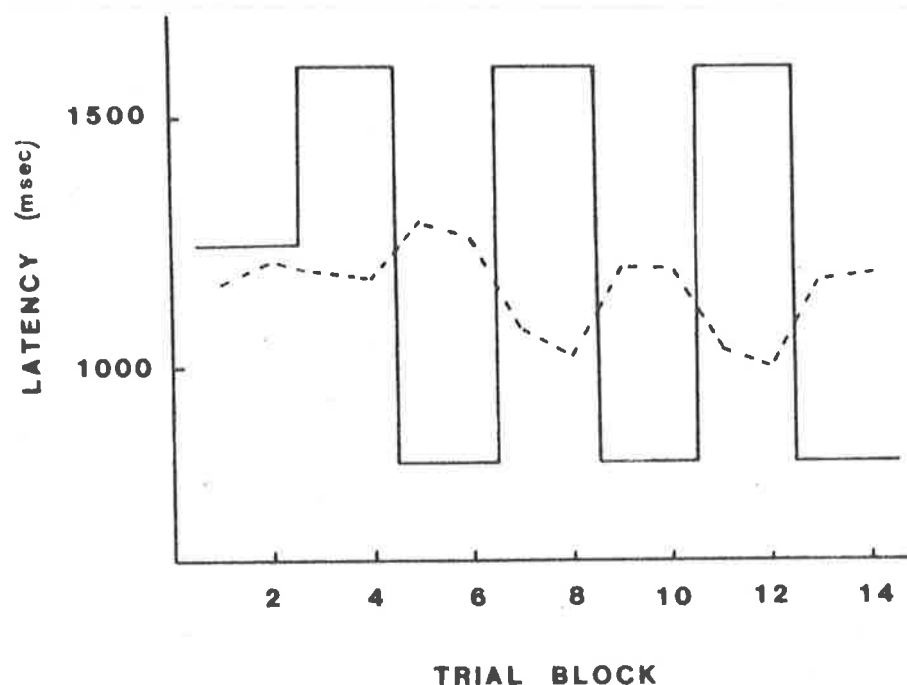


FIGURE 79 The mean latency of all responses across the line length differences in each block of 48 trials. The unscaled level of discriminability is given by the unbroken line.

in each block of trials. In agreement with both models, response times tended to be shorter at high discriminability, while latency appeared to show a general downward trend during the task. A tendency for latency to decrease at high discriminability became more marked over the final two sequences of high discriminability, while a similar tendency for latency to decrease at low discriminability seemed to disappear over the final two sequences of low discriminability.

Table 33 gives the results of analyses of variance on correct and incorrect response latencies, as well as the combined latency of all responses, performed on the data from all shifts in discriminability. The main effect of discriminability was significant for both correct and overall response latencies. These two measures also yielded significant period effects, but no interaction, so that latencies generally tended to decline during the

TABLE 33. Results of analyses of variance on response latencies averaged across the three line length differences in each block of trials.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Correct response</u>				
	Discriminability (D)	1	4.534	37.265	.00001
	D x S	27	.121		
	Period (P)	1	.083	8.124	.01
	P x S	27	.010		
	D x P	1	.029	1.980	
	D x P x S	27	.014		
(b)	<u>Incorrect response</u>				
	Discriminability (D)	1	3.250	2.232	
	D x S	27	1.455		
	Period (P)	1	.146	.814	
	P x S	27	.179		
	D x P	1	1.025	1.199	
	D x P x S	27	.854		
(c)	<u>Overall</u>				
	Discriminability (D)	1	5.427	37.746	.00001
	D x S	27	.143		
	Period (P)	1	.099	8.184	.01
	P x S	27	.012		
	D x P	1	.056	3.271	
	D x P x S	27	.017		

sequence following each step change in discriminability, irrespective of its direction. No effects were found on incorrect response latencies, but these responses were relatively few, especially, as indicated in Figure 78, during the sequences of high discriminability. During the task, a decreased tendency for latencies to decline at low discriminability was indicated when further analyses were performed which excluded data from the first two shifts in discriminability. A significant interaction between discriminability and period then emerged in both correct and overall response latencies ($F = 6.525$, $d.f. = 1/27$, $p < .05$, and $F = 9.134$, $d.f. = 1/27$, $p < .01$, respectively). In both cases simple period effects indicated that this was due to significant decreases in response times over sequences of high discriminability ($F = 9.774$, $d.f. = 1/27$, $p < .005$, and $F = 22.183$, $p < .0001$, respectively), while there

was no change over sequences of low discriminability ($F < 1.0$ in both cases).

Although the decrease in latency at high discriminability is consistent with the decrease in caution specified by the accumulator model, it was not accompanied by a complementary increase in latency over sequences of low discriminability. However, any such tendency for latency to increase may have been obscured by the general downward trend in response latency that appears to be superimposed on the performance curve of Figure 79, which results in a downward modulation with time. Comparisons of overall response latency showed significant declines from the first to final sequences of high discriminability ($F = 39.066$, d.f. = 1/27, $p < .00001$), as well as the first to final sequences of low discriminability ($F = 10.692$, d.f. = 1/27, $p < .005$), which gives evidence of such a general decline in latency. Similar analyses revealed that, while correct response probability increased from the first to final sequence of high discriminability ($F = 4.386$, d.f. = 1/27, $p < .05$), there was no change from the first to final sequence of low discriminability ($F < 1.0$). Thus, the general decline in response latency may, or may not, have indicated a general increase in task proficiency with time.

(iii) Response confidence. The trends shown in response confidence differed for correct and incorrect responses. Figure 80 shows that correct response confidence was higher at high discriminability, that it fell during sequences of high discriminability, and that it showed a (lesser) tendency to increase during sequences of low discriminability. These trends are consistent with both models, since the greater separation between sensory distributions at high discriminability will tend to increase confidence in decisions above target levels, so that adjustments are required to reduce response confidence. Meanwhile, at low discriminability adjustments will

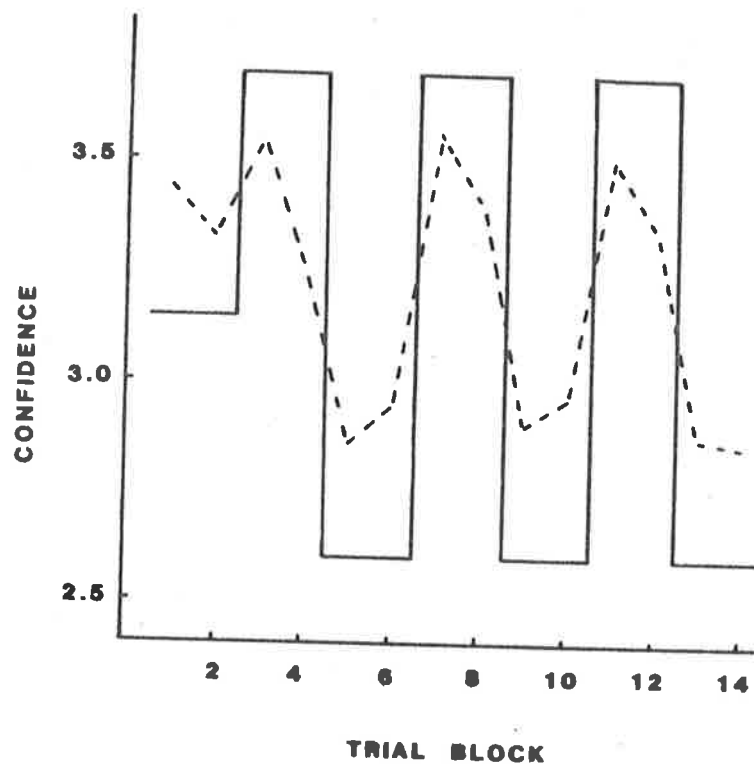


FIGURE 80 The mean confidence expressed in correct responses across the line length differences in each block of 48 trials. The unscaled level of discriminability is given by the unbroken line.

be required to increase confidence toward its target level. Due to the small weighting on overall confidence of the relatively infrequent incorrect responses, the same trends were also shown in confidence averaged across all responses in each block of trials.

Table 34 confirms the main effect of discriminability on both correct response confidence and overall confidence. The period effect was also significant on both measures, and can be attributed to the decreases in confidence, during the sequences which followed a step increase in discriminability, which were larger than the increases in the sequences which followed a step decrease. As a result, an average downward trend is indicated after each step change in discriminability. This was confirmed by tests of the

TABLE 34. Results of analyses of variance on response confidence averaged across the three line length differences in each block of trials.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Correct response</u>				
	Discriminability (D)	1	73.628	183.983	.00001
	D x S	27	.394		
	Period (P)	1	1.434	6.417	.05
	P x S	27	.223		
	D x P	1	4.106	22.737	.0001
	D x P x S	27	.180		
(b)	<u>Incorrect response</u>				
	Discriminability (D)	1	4.518	3.501	
	D x S	27	1.290		
	Period (P)	1	.070	.083	
	P x S	27	.845		
	D x P	1	2.442	1.754	
	D x P x S	27	1.392		
(c)	<u>Overall</u>				
	Discriminability (D)	1	77.766	177.074	.00001
	D x S	27	.439		
	Period (P)	1	1.041	5.008	.05
	P x S	27	.208		
	D x P	1	3.392	19.831	.0001
	D x P x S	27	.171		

interaction effects which accompanied each pair of main effects. Both correct response confidence and overall confidence fell significantly during sequences of high discriminability ($F = 24.870$, $d.f. = 1/27$, $p < .00005$, and $F = 16.642$, $d.f. = 1/27$, $p < .0005$, respectively), but increased to a nonsignificant extent during sequences of low discriminability ($F = 1.758$, $d.f. = 1/27$, $p = .20$, and $F = 2.536$, $d.f. = 1/27$, $p = .12$, respectively). Thus, as was found with response latencies, change was more pronounced during the sequence of trials following a step increase in discriminability than the sequence following a step decrease. Unlike latency and accuracy, there was no general tendency for confidence to increase or decrease during the

task, with comparisons of overall confidence showing no change from the first to final sequence of either high or low discriminability ($p > .40$ in all cases).

Figure 81 shows that confidence in incorrect responses also tended to rise during sequences of low discriminability and to fall during sequences of high discriminability, but that the former rise was from an already increased level of confidence, and the latter fall from an already lowered level of confidence. That is, the discriminability effect was reversed for incorrect responses, with confidence being lower during sequences of high discriminability. However, the main effect of discriminability reported in Table 34 was only of borderline significance ($p = .07$). Nevertheless, this was rather strong, considering the comparatively small pool of data upon

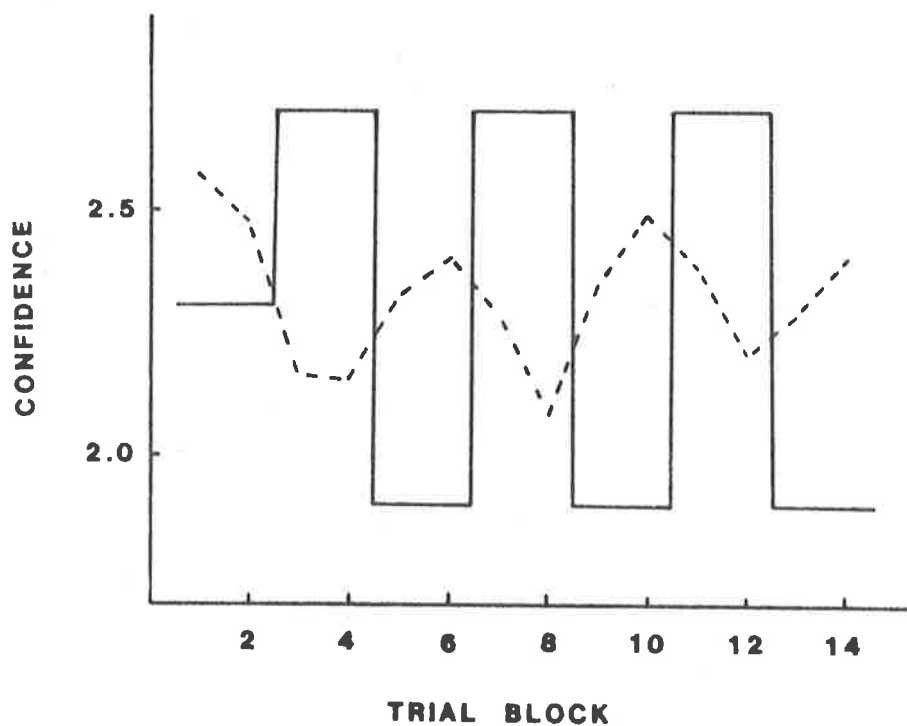


FIGURE 81 The mean confidence expressed in incorrect responses across the line length differences in each block of 48 trials. The unscaled level of discriminability is given by the unbroken line.

which the analysis was performed. A check of the raw data indicated that incorrect response confidence was higher during sequences of low discriminability than during sequences of high discriminability for 21 of the 28 observers. The binomial probability, under a 2-tailed hypothesis, that this occurred due to chance is .013. The relative infrequency of incorrect responses may also have been responsible for the lack of a significant interaction on incorrect response confidence (Table 34). Figure 81 strongly suggests the same trends in the confidence of incorrect responses during sequences of high and low discriminability as those shown in Figure 80 for correct responses.

(b) Performance at line length difference 3

When performance had been averaged across line length differences, adjustments in response to step changes in discriminability were inferred by the trends which occurred during sequences of high and, to a lesser extent, low discriminability. The analyses to be reported in the present section sought to determine whether there had been variation during the task in responses to a constant line length difference and, in so doing, to obtain direct evidence that decision criteria had been regulated.

(i) Response probability. Change during the session in the probability of correctly reporting a stimulus of line length difference 3 was the same whether the stimulus had longer lines to the right or to the left, and is shown in Figure 82 for left and right responses combined. Variation was definitely indicated in the measure and, as in the case of accuracy averaged across line length differences, there is a discontinuity when the trends shown over the first two shifts in discriminability are compared with those over the final four shifts. Accuracy was higher in the first sequence of low discrimin-

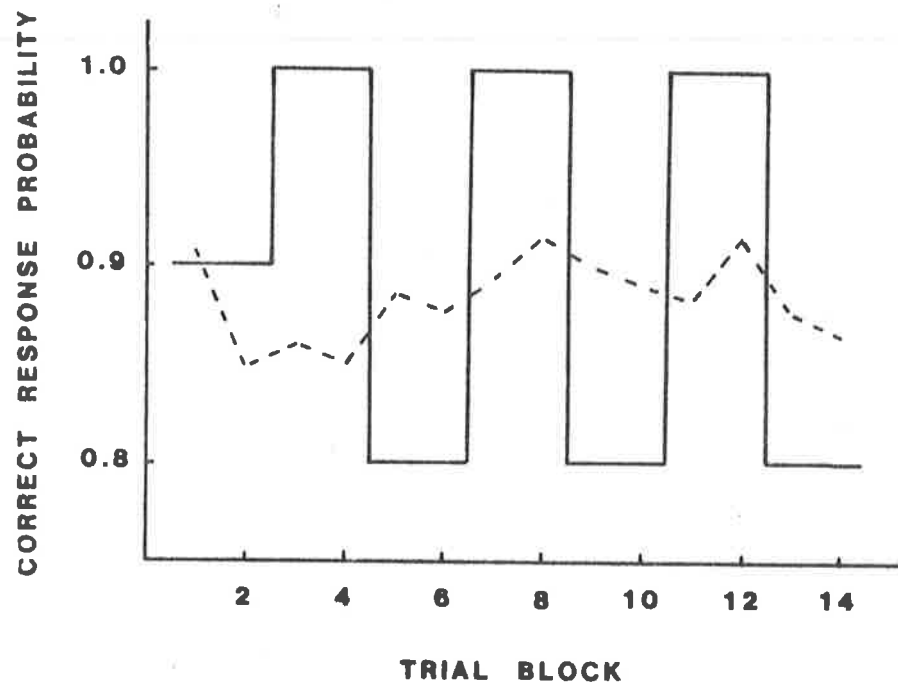


FIGURE 82 The mean probability of making a correct response to stimuli with a line length difference of 3 in each block of 48 trials. The unscaled level of discriminability is given by the unbroken line.

ability than in the preceding sequence of high discriminability, and tended to decline within both sequences. Thereafter, the probability of a correct decision did not appear to be related to discriminability level, but, instead, consistently increased during the sequences of high discriminability and decreased during sequences of low discriminability. As a result of this dichotomy, separate analyses were again performed on the data from the first two shifts in discriminability and those from the final four shifts in discriminability.

As reported in Table 35, there were no significant effects on correct response probability over the first two shifts in discriminability. Over the final four shifts, however, there was a significant interaction between discriminability level and period. Tests of this interaction found it to be the result of a significant increase in correct response probability during

TABLE 35. Results of analyses of variance on the probability of making a correct report of a stimulus with a line length difference of 3.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Over trial blocks 3 to 6</u>				
	Discriminability (D)	1	.018	2.014	
	D x S	27	.009		
	Period (P)	1	.002	.382	
	P x S	27	.006		
	D x P	1	.000	.003	
	D x P x S	27	.008		
(b)	<u>Over trial blocks 7 to 14</u>				
	Discriminability (D)	1	.019	3.326	
	D x S	27	.005		
	Period (P)	1	.003	1.044	
	P x S	27	.003		
	D x P	1	.020	4.615	.05
	D x P x S	27	.004		

sequences of high discriminability ($F = 7.010$, $d.f. = 1/27$, $p = .05$), and no significant change during sequences of low discriminability ($F = .621$, $d.f. = 1/27$, $p > .4$). Although there was little or no effect at low discriminability, the increase in accuracy during later high discriminability sequences agrees with the increase in caution predicted by the dual-criterion model.

Some indication that responses became generally more accurate during the session was given by a significant increase in correct response probability from the first to final sequence of high discriminability ($F = 6.986$, $d.f. = 1/27$, $p = .01$), although there was no significant change from the first to final sequence of low discriminability ($F = .487$, $d.f. = 1/27$, $p > .45$).

(ii) Response latency. Due to the high degree of accuracy of responses to stimuli of line length difference 3 (see Figure 82), too few errors were committed to allow reliable statistical analyses of change during the session in incorrect response latencies. However, the omission of a separate analysis of incorrect response latencies does not neglect any feature

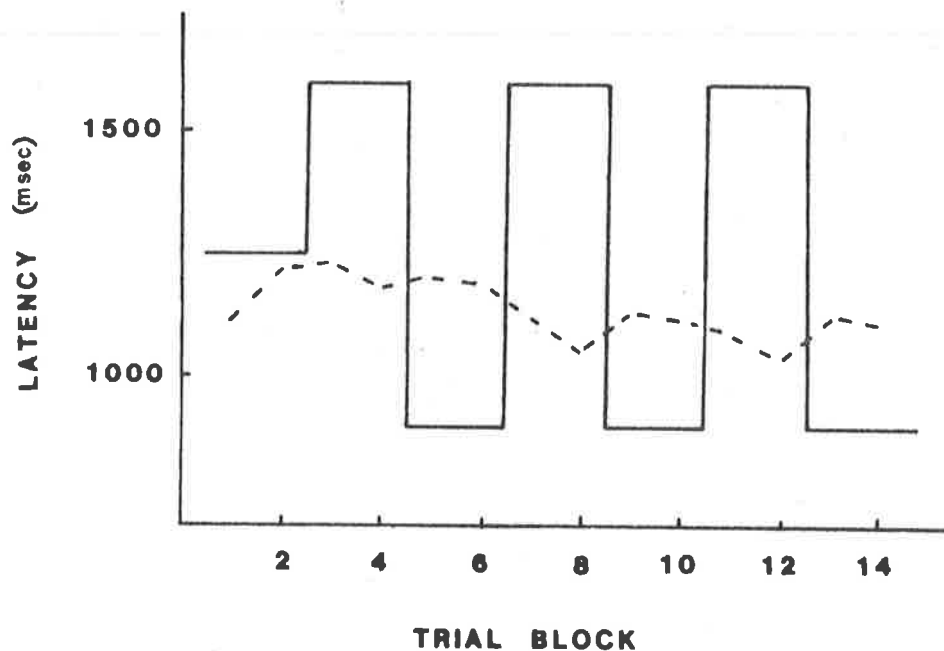


FIGURE 83 The mean latency of all responses to stimuli with a line length difference of 3 in each block of 48 trials. The unscaled level of discriminability is given by the unbroken line.

of the data since the same trends were shown in the latencies of correct and incorrect responses to both types of stimulus, and are shown in Figure 83 for the mean latency of all responses. Superimposed on a general tendency for latencies to decrease during the session was a further tendency for latency to be higher during sequences of low discriminability, as would occur if primary decision criteria, and hence caution, were higher during these sequences, in line with the predictions of the accumulator model. Latency again tended to decline during the sequence which followed each step change in discriminability, but mainly during those which followed a step increase. This latter period effect is confirmed in Table 36 for both correct response latency as well as the latency of all responses combined. The tendency for latency to be higher at low discriminability is also confirmed by the significant effect of discriminability level on overall latency and a marginal effect ($p = .072$) on correct response latency. Tests

TABLE 36. Results of analyses of variance on the latencies of responses to stimuli with a line length difference of 3.

	<u>Source</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Correct response</u>				
	Discriminability (D)	1	.056	3.498	
	D x S	27	.016		
	Period (P)	1	.092	11.421	.005
	P x S	27	.008		
	D x P	1	.007	.813	
	D x P x S	27	.009		
(b)	<u>Overall</u>				
	Discriminability (D)	1	.065	4.222	.05
	D x S	27	.015		
	Period (P)	1	.103	14.160	.001
	P x S	27	.007		
	D x P	1	.035	4.126	.05
	D x P x S	27	.008		

of the interaction effect on overall latency showed it to be due to significant declines in latency during sequences of high discriminability ($F = 16.225$, $d.f. = 1/27$, $p < .0005$) and no change during sequences of low discriminability ($F = 1.137$, $d.f. = 1/27$, $p > .25$). Hence the period effect resulted mainly from change during sequences of high discriminability.

The general decline in response latencies was indicated by a significant decrease in overall latency between the first and final sequence of high discriminability ($F = 21.722$, $d.f. = 1/27$, $p < .0001$) as well as between the first and final sequence of low discriminability ($F = 5.715$, $d.f. = 1/27$, $p < .05$). This trend could be responsible for the greater declines in latency found during sequences of high discriminability than sequences of low discriminability since it would have accentuated the main effect of discriminability on latency in the former case and attenuated it in the latter.

(iii) Response confidence. The trends which appeared in response confidence were the same for responses to both types of stimulus and, unlike the case where confidence had been averaged across stimulus differences, were also the same for correct and incorrect responses. As shown in Figure 84, response confidence appeared to display a more orderly pattern over the final four shifts in discriminability than it had shown over the first two shifts. Specifically, while confidence was lower during the first sequence of high discriminability than the first sequence of low, it tended to decline during each sequence. From then on, however, confidence always declined at high discriminability and increased at low discriminability, with the former decrease being the larger so that confidence appeared to be on average higher during sequence of high discriminability. For this reason,

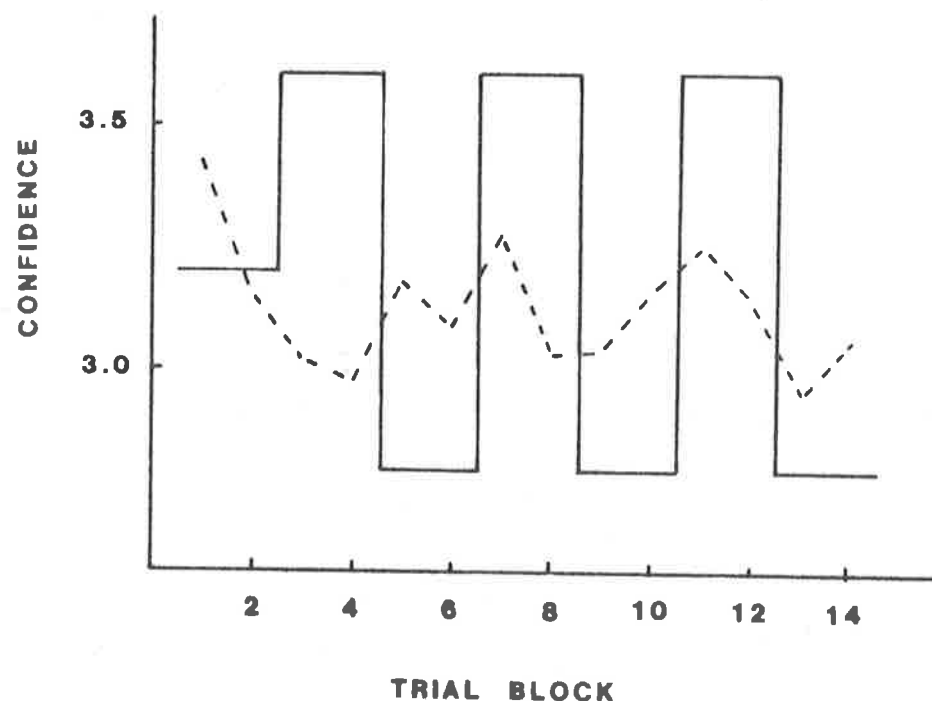


FIGURE 84 The mean confidence expressed in all responses to stimuli with a line length difference of 3 in each block of 48 trials. The unscaled level of discriminability is given by the unbroken line.

separate analyses were performed on the data from the first two shifts in discriminability and those from the final four shifts. As had been found with response latencies, no reliable analyses could be performed on the confidence of incorrect responses.

There was no significant main effect or interaction on either correct response confidence or overall confidence during the first two shifts in discriminability ($p > .10$ in all cases). However, the greater consistency indicated in these measures over the final four shifts in discriminability led to the significant effects reported in Table 37 that were obtained over this latter part of the session. The main effect of discriminability level was significant on overall response confidence and marginal ($p = .057$) on correct response confidence, and so confirms that confidence tended to be generally higher during sequences of high discriminability. Significant interactions also appeared on both correct response confidence and overall confidence,

TABLE 37. Results of analyses of variance on the confidence of responses to stimuli with a line length difference of 3 over trial blocks 7 to 14.

	<u>Source</u>	<u>d.f</u>	<u>Mean Square</u>	<u>F</u>	<u>Significance</u>
(a)	<u>Correct response</u>				
	Discriminability (D)	1	.672	3.926	
	D x S	27	.171		
	Period (P)	1	.147	1.568	
	P x S	27	.093		
	D x P	1	1.247	10.929	.005
	D x P x S	27	.114		
(b)	<u>Overall</u>				
	Discriminability (D)	1	.819	5.950	.05
	D x S	27	.137		
	Period (P)	1	.057	.566	
	P x S	27	.101		
	D x P	1	1.264	11.678	.005
	D x P x S	27	.108		

and in each case was due to a significant decline in confidence during sequences of high discriminability ($F = 7.849$, $d.f. = 1/27$, $p < .01$, and 5.929 , $d.f. = 1/27$, $p < .05$, respectively) and a significant increase during sequences of low discriminability ($F = 4.175$, $d.f. = 1/27$, $p = .05$, and $F = 7.447$, $d.f. = 1/27$, $p = .01$, respectively). These latter simple effects of period corroborate the trends found to follow a step change in discriminability when confidence was averaged across line length differences. Further, the effects are here significant during sequences of both high and low discriminability and thus confirm the predictions of both models with regard to change in confidence in response to a step change in discriminability. It should be noted that the main effect of discriminability, higher confidence at high discriminability, appears at first glance to run counter to the predictions of both models. However, Figure 84 indicates that this effect probably resulted from tendencies over the latter part of the session for the increase in confidence at low discriminability to continue on into the first part of the following sequence of high discriminability, and for the decrease at high discriminability to continue into the following sequence of low discriminability.

Corresponding to those tendencies indicated during the session for responses to become faster, and, at least in comparisons of sequences of high discriminability, more accurate, the confidence in all responses to stimuli of line length difference 3 was higher in the last sequence of high discriminability than in the first ($F = 6.860$, $d.f. = 1/27$, $p = .01$) although there was no change from the first to final sequence of low discriminability ($F = 1.976$, $d.f. = 1/27$, $p > .15$).

(c) Early trends in performance

Data from the first two trial blocks of the task, in which stimuli were presented at the same median level of discriminability as in practice, were purposely omitted from the previous analyses. However the general trends shown for responses to be made more accurately, quickly, and confidently, as time progressed, suggests that there may have been an increase in proficiency unrelated to the experimental manipulation of stimulus discriminability. Therefore, the data from the first two trial blocks were analysed to see whether there had been any trend toward performance improvement at this early stage of the task.

Table 38 gives the mean measures of performance in trial blocks 1 and 2 for responses made to all stimuli as well as those to stimuli of line length difference 3. From this table it can be seen that, in contrast to these trends observed later in the task, there appeared to be a general decrement in performance between the first and second block of trials.

TABLE 38. Mean performance measures during trial blocks 1 and 2 (median discriminability).

<u>Response measure</u>	<u>(a) Across all differences</u>		<u>(b) At line length difference 3</u>	
	<u>Trial block</u>		<u>Trial block</u>	
	1	2	1	2
Correct response probability	.82	.81	.91	.85
Correct response latency (msec)	1121	1193	1106	1203
Incorrect response latency (msec)	1206	1308	1114	1289
Overall latency (msec)	1130	1202	1111	1218
Correct response confidence	3.4	3.3	3.5	3.3
Incorrect response confidence	2.6	2.5	2.6	2.2
Overall confidence	3.4	3.2	3.4	3.2

Analyses of variance performed on the data from all responses showed the change in correct response probability to be nonsignificant ($F = .687$, d.f. = $1/27$, $p > .40$). However, the increase in overall response latency and decrease in overall confidence were both significant ($F = 11.749$, d.f. = $1/27$, $p < .005$, and $F = 4.728$, d.f. = $1/27$, $p < .05$, respectively), and, in each of these cases, there was also significant change in the measure based on correct responses ($F = 11.163$, d.f. = $1/27$, $p < .005$, and $F = 4.033$, d.f. = $1/27$, $p = .05$, respectively), but no change in the measure for incorrect responses ($F = 2.066$, d.f. = $1/27$, $p > .15$, and $F = .522$, d.f. = $1/27$, $p > .50$, respectively).

When only the data arising from responses to stimuli of line length difference 3 were analyzed, the decrease in correct response probability from the first to the second trial block was significant ($F = 8.286$, d.f. = $1/27$, $p < .01$). The associated increase in overall response latency was significant ($F = 25.690$, d.f. = $1/27$, $p < .00005$). Again, while the increase in correct response latency was also significant ($F = 21.506$, d.f. = $1/27$, $p < .0001$), the change in incorrect response latency was not ($F = 1.693$, d.f. = $1/27$, $p > .20$). Similarly, the decrease in overall response confidence was significant ($F = 9.299$, d.f. = $1/27$, $p = .005$), as was the decrease in the confidence of correct responses ($F = 5.412$, d.f. = $1/27$, $p < .05$), but not the change in incorrect response confidence ($F = .192$, d.f. = $1/27$, $p > .65$). Thus, the general trends observed as discriminability alternated later in the session cannot be considered as a continuation of any earlier tendency for performance to improve. If anything, these later trends had cancelled a previous tendency toward a performance decrement.

3. Discussion

(a) The effects of variation in caution - an anomaly

The step changes in discriminability had the desired effect of producing more pronounced variation in performance than had previously been observed to follow a ramp change in discriminability. As a consequence, there was now clear evidence that observers had responded to the changes which were produced in the characteristics of their ongoing performance. In agreement with both models, it appeared that adjustments occurred which sought to maintain some target level of confidence in responses. Whenever the confidence expressed in responses had been increased by a step increase in discriminability, or decreased by a step decrease, there followed a period during which confidence tended to return toward some intermediate level. Such trends were indicated in averaged performance during each sequence which followed a step change in discriminability, as well as in performance at a constant amount of stimulus difference, the latter indicating that decision criteria were being adjusted in order to compensate for the variation in response confidence. However, while both models agree that compensation will occur through adjustments in the degree of caution exercised by the observer in his responses, the patterns exhibited by response latency and accuracy do not both conform to the relationship that would be predicted by either model.

The inverse relationship between degree of caution and level of discriminability that is specified by the accumulator model was strongly supported by the finding that the latencies of responses to the constant amount of stimulus difference were longer at low levels of discriminability. Both theoretical viewpoints agree that this indicates a higher degree of caution, since it is only through variation of decision criteria, which directly

determine the number of observations which must be sampled prior to a decision, that the observer can produce systematic changes in response accuracy and confidence. However, while confidence increased during sequences of low discriminability, accuracy did not also increase in the expected manner. Although accuracy had shown an early tendency to increase at low discriminability and decrease at high, when it was averaged across all stimulus differences, and was inversely related to discriminability level when measured at the constant stimulus difference, none of these trends was strong enough to reach significance. Indeed, over the latter part of the session, the significant trends were for accuracy to decrease at low discriminability and increase at high. Thus, paradoxically, the higher latencies at low discriminability resulted in decreased, rather than increased, accuracy, and vice versa for the lower latencies at high discriminability.

It ought to be mentioned that the dual-criterion detection model can be altered so that it too predicts the inverse relationship between caution and discriminability that is implied by the latency data. This can be accomplished if it is assumed that confidence is based solely on the sensory magnitude of an observation, rather than the distance that the observation lies from the interval of uncertainty (IU). For example, confidence might increase monotonically from either side of that sensory value which occurs at the intersection of the two sensory distributions, as shown in Figure 85, and this functional relationship will remain invariant as the IU expands or contracts. Then, as is also assumed by a distance-from-criterion function, average confidence will decrease as the distributions move closer together during a decline in discriminability. However, in the present model, confidence will subsequently be increased by expanding the IU in order to increase the proportion of observations associated with higher levels of confidence.

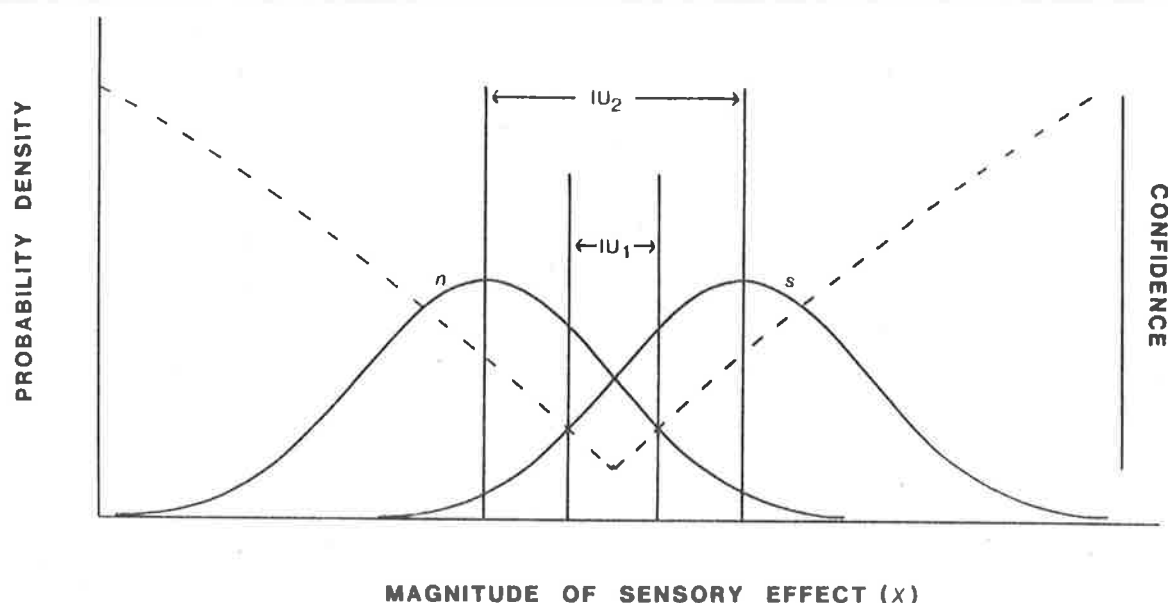


FIGURE 85 A dual-criterion detection model in which confidence, given by the dashed line, is assumed to be a function solely of sensory magnitude. As the interval of uncertainty expands from IU_1 to IU_2 , there is an increase in the average confidence associated with values of x that lie outside the interval.

Thus, confidence regulation will produce an inverse relationship between caution and the level of discriminability, as also occurs in the accumulator model and was indicated by the latency and confidence data.

By reversing the relationship between caution and discriminability, the dual-criterion model negates the favourable evidence that the accuracy data provides for the predictions of the original detection model. This follows because the same direct relationship still exists between accuracy and latency as the width of the IU is varied. An increase in latency through expanding the IU at low discriminability should then produce more accurate responses, rather than the trends toward decreased accuracy which were found later in the experimental session. Hence, the model is confronted with the same dilemma as the accumulator model : the weaker trends in

accuracy over the early part of the session conform to the predictions of the model, while the stronger trends that were later shown are totally contradictory. In order for any model to explain the present set of results, it must have the capacity to account for an inverse relationship between response accuracy and latency, as well as the direct relationship that exists between them in the speed-accuracy tradeoff normally assumed under conditions of varying caution.

(b) A parallel-control model

It is possible that the present results stemmed from the action of two independent regulatory systems, only one of which is directly involved with the control of the observer's decision process, but the second of which might indirectly affect the accuracy and speed of decisions. Specifically, it could be that decision criteria are controlled by a confidence regulator like that specified by Vickers (1978, 1979), while the rate at which sensory data enter the primary decision process is simultaneously controlled by a second system which operates according to the basic principles specified in Pike's interval of uncertainty model (Pike, McFarland & Dalglish, 1974; Pike & Dalglish, 1976). A configuration of two such processes working in parallel is shown in Figure 86. According to this model there is an interval of sensory indifference (ISI) in which the magnitudes of sensations provide too little evidence to allow classification according to the available response alternatives. These magnitudes are those which lie close to a likelihood ratio of 1.0, in a signal detection model, or the point of sensory indifference, in the accumulator model. When one of these magnitudes is sampled, rather than adding it to one of the accumulated totals in the decision process, a further observation is made. Observations will thus

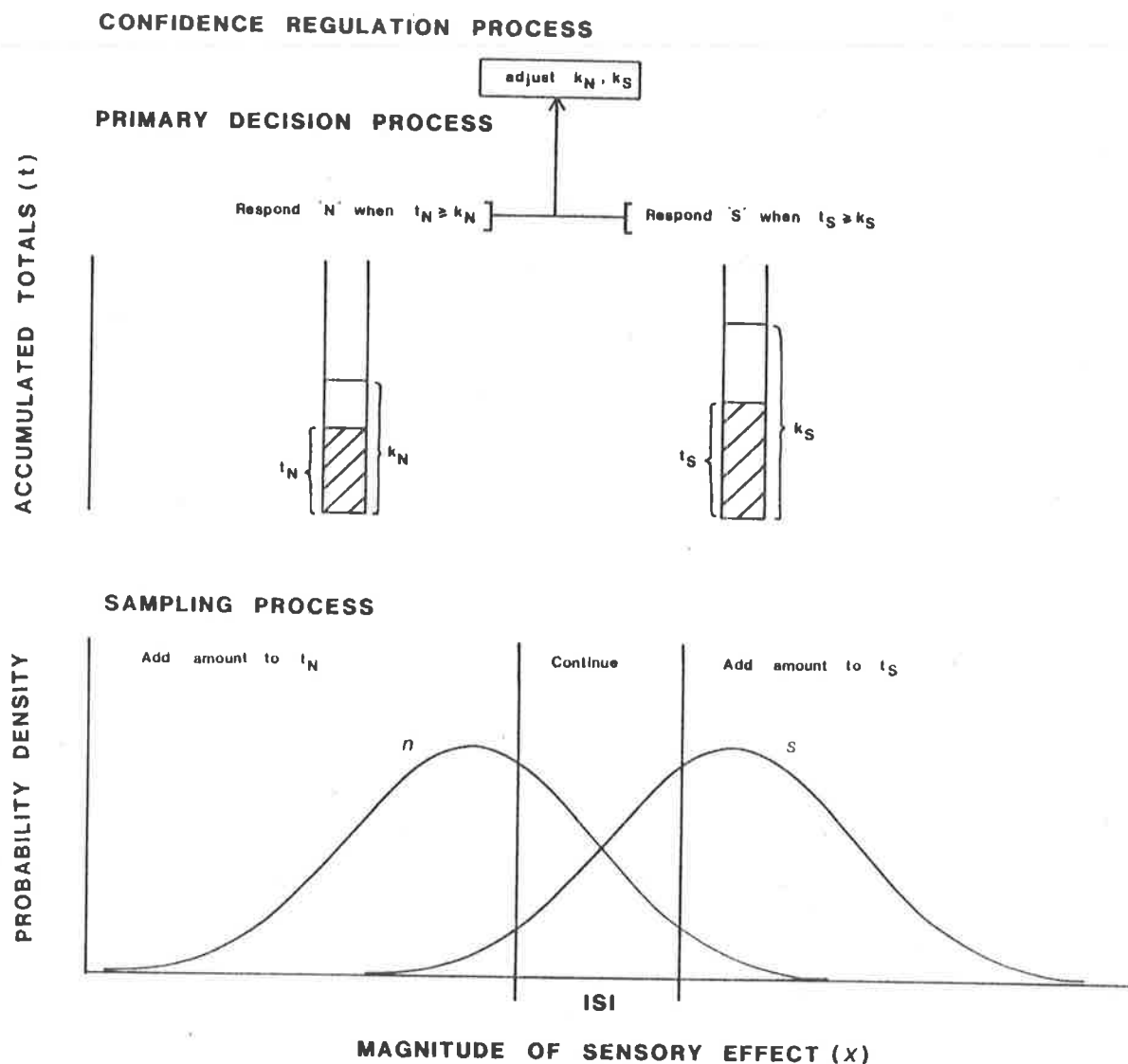


FIGURE 86 Decision and sampling processes in the parallel-control model. An observation, x , sampled from the interval of sensory indifference, ISI, provides evidence for neither decision alternative. When x is sampled outside the ISI an amount proportional to the distance from x to the ISI is added to t_N , if x falls below the ISI, or t_S , if x exceeds the ISI. Observations proceed until either t_N or t_S , respectively, has reached its critical total, k_N , or k_S , whereupon a nonsignal (N) or signal (S) response is initiated. The value of k_N and k_S is each varied according to a confidence regulation process like that shown in figure 69.

proceed until the evidence that has been sampled outside the ISI has reached a critical total for one of the response alternatives, whereupon that response is initiated. The confidence arising from that response will then be used to determine whether any adjustment is necessary to the value of the response criterion. The entire process is depicted in Figure 86 for the case of a signal detection task.

The main departure of the model from the accumulator model is that it proposes the existence of 'empty' observations, that is, intervals in the sampling process during which evidence is not passed on to either accumulator, but which can be assumed to have occupied the same finite period of time as those intervals during which some information from the sampling process does enter the decision process. A characteristic of this sort of sampling process which would require regulation is the rate at which it allows information to enter the decision process. Since this rate is determined by the proportion of empty observations, control is achieved by adjusting the positions of the two sensory boundaries which limit the width of the ISI, and thereby altering the proportion of observations which lie in the ISI. The goal of such regulation would conceivably be to maintain, at some constant value, the rate at which information is processed by the decision process, the particular rate being determined by the constraints placed on the decision process due to the nature, or quality, of the information it receives.

While it can be assumed that the degree of emphasis placed on the speed or accuracy of performance will normally only influence the target levels of confidence that are sought in the control accumulator process, there is a considerable tradeoff in the sampling process between the speed of information processing and the accuracy of the information that is

processed. Figure 87 illustrates this for the example of an ISI that expands from either side of the intersection of two sensory distributions whose means are separated by 2.0 standard deviations. As the proportion of empty observations here increases to almost .7, there is a decline in the average distance from the ISI of those observations which provide evidence for the correct response alternative as well as those which provide evidence for the incorrect response alternative. Thus there is a double cost of

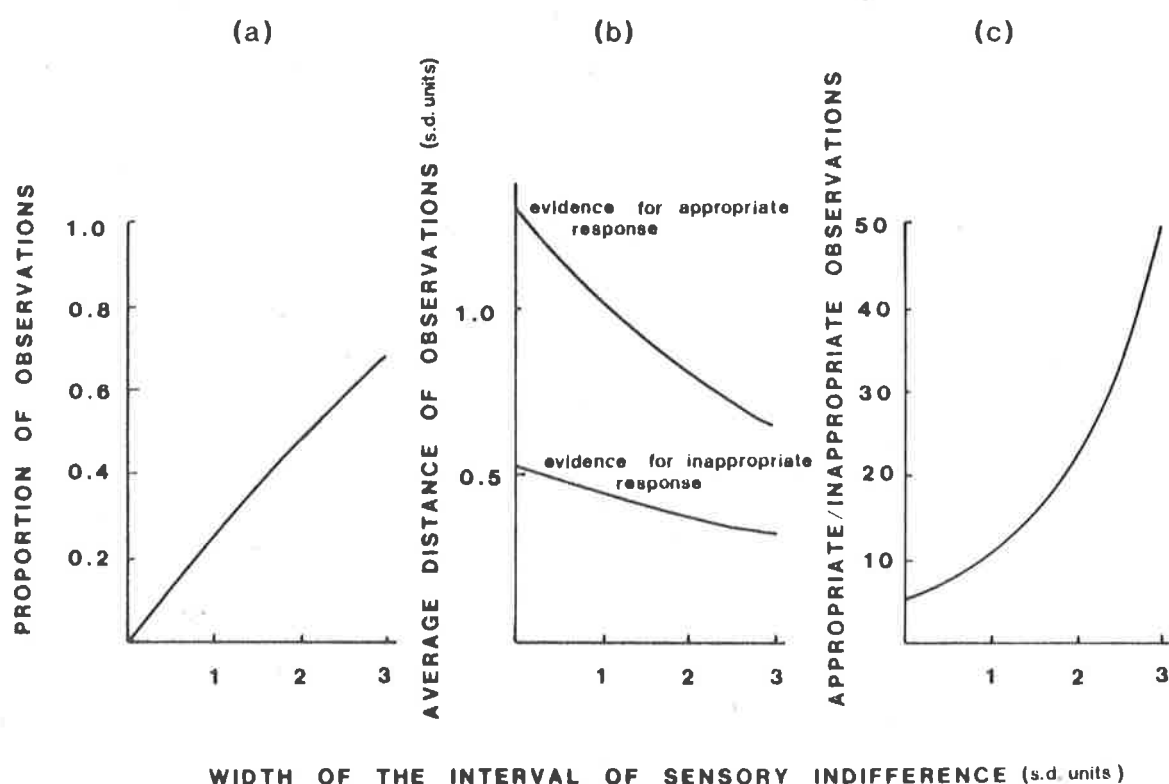


FIGURE 87 Relationship between the width of the interval of sensory indifference, ISI, and (a) the proportion of observations lying within the interval, (b) the average distance from the interval of observations that provide evidence for an appropriate response and those which provide evidence for an inappropriate response, and (c) the ratio of the proportion of observations which provide evidence for an appropriate response to the proportion of those which provide evidence for an inappropriate response. The ISI is symmetrically located about the intersection of two equal-variance sensory distributions whose means are separated by 2.0 standard deviations.

expanding the ISI : as well as increasing the proportion of observations which provide evidence for neither decision alternative, there is a decrease in the average amount of evidence which does enter the decision process, the latter assuming a distance-from-ISI function for sensory evidence.

For any given criterion level, both of these effects will increase the time taken to reach a decision. However, these costs on response latency could be well outweighed by the marked benefit to the accuracy of responses which is also shown in Figure 87. As the proportion of empty observations increases from 0.0 to .7, there is a tenfold increase in the ratio of observations which provide evidence in favour of the correct response to those which favour the incorrect response. Thus, to give the simplest example, that of an observer who has adopted such a low criterion that a response will be based on a single observation outside the ISI, the expansion of the ISI from 0.0 to 3.0 standard deviations has, according to equations 17 and 18, resulted in 2.2 more observations being required, on average, to reach a decision, but the probability that the decision is correct has increased from .84 to .98.

It seems reasonable to assume that once the width of the ISI has been established for a given experimental situation, the regulation of the sampling process ensures that the sensory boundaries of the ISI will be maintained at fairly constant values. Under such circumstances, the characteristics of responses will be affected only by the adjustments made to decision criteria in the regulation of response confidence by the control accumulator. However, responses will be affected by the sampling regulation if there is change at the sensory level which affects the sampling rate in such a way that the ISI has to be expanded or contracted. Such a circumstance presented itself in the last experiment, since, for a given

ISI width, the increased overlap of sensory distributions at low discriminability results in a higher proportion of empty observations than occurs when discriminability is high. Therefore, in order to produce a sampling rate common to both situations, the ISI will have to be made narrower when discriminability is low, and wider when it is high. As a result, whenever there was a step increase in discriminability, the effects of an expanding ISI would have occurred in conjunction with the decreases in decision criteria demanded by the control accumulator. Conversely, at low discriminability, the effects of a contracting ISI would have accompanied increases in decision criteria.

The two processes operating at any time in the task would have had antagonistic effects on measures of performance. For instance, at high discriminability the decrease in criterion levels demanded by the control accumulator would decrease the amount of evidence that has to be sampled prior to a response so that, on its own, this would decrease the accuracy of responses as well as their latency. If the ISI is expanding at the same time, however, the trend toward decreasing accuracy could be cancelled and, if the expansion is great enough, would be reversed by a manyfold increase in the proportion of observations which provide evidence for the correct response, relative to those which favour the incorrect one. At the same time the tendency for response criteria to fall would counter the increased sampling time that usually accompanies this increased accuracy. In particular, this would occur if criteria were lowered far enough so that, by the end of the expansion of the ISI, responses would often be elicited by the first observation sampled outside the ISI. In such circumstances, the average number of observations, filled and empty, that are sampled prior to a response when the ISI is wide will be less than the number which has to be sampled when the ISI is narrow, but decision criteria higher. For

instance, to return to the example given in Figure 87, as the width of the ISI increases from 0.0 to 3.0 standard deviations, the expected number of observations which occur before one is sampled outside the ISI increases from 1.0 to 3.2. Thus, in this example, in order for an actual decline in latency to occur while the ISI expands, decision criteria will only have to change to a degree that allows a response to be based, on average, on 2.3 fewer favourable observations at the end of the expansion than were required at the beginning.

The opposite pair of antagonist effects, those produced by decision criteria which are being raised while the ISI is being narrowed, could explain the increases in latency, but decreased accuracy, at low discriminability. However, the variation in the ISI would not have interfered with the effects that changes in caution had on response confidence, since this measure is most directly influenced by the values of decision criteria. For instance, even though the evidence being sampled at high discriminability became more favourable to the correct response alternative, confidence, as measured by the balance of evidence between accumulated totals, would have tended to fall as the decline in decision criteria brought about a decrease in the maximum possible balance which could be attained. Moreover, an expansion of the ISI leads to a decrease in the average size of the amounts of evidence that are accumulated, as well as an increase in the likelihood that each amount will favour the correct response (Figure 87), thereby allowing more accurate decisions to be based on smaller balances of evidence in their favour. Thus, decision criteria which decline concurrently with an expanding ISI will provide the necessary conditions for lower confidence to follow from more accurate responses.

The degree to which accuracy is improved at the expense of slowing down the sampling rate is an indirect function of the level of discriminability of stimuli. This is illustrated in Figure 88 which gives the ratio of observations which provide evidence for a correct response to those which provide evidence for an incorrect response, and the average minimum number of observations which occur before one is sampled outside the ISI, for a symmetrically located ISI that expands from 1.0 to 3.0 standard deviations, and sensory distributions whose means are separated by 1.0, 2.0, or 3.0 standard deviations. As is evident in the figure, the rate at which accuracy increases is greatest when the ISI expands at high discriminability, while the rate at which sampling rate decreases is the least. Thus the degree to which an expansion of the ISI will counter the effects on performance of a decrease in caution at high discriminability will be greatest for the accuracy measure and least for latency. Conversely, accuracy will be less affected at low discriminability, and latency more affected, by a contraction of the ISI. In support of this, with performance measured across all stimulus differences, the increases in accuracy at high discriminability (interpreted as the effect of sensory regulation rather than change in caution) reached a level of significance that was ten times greater than that reached by the complementary decreases at low discriminability (.005 as opposed to .05). Further, the decreases in latency at high discriminability (interpreted as the effect of a decrease in caution) were significant, whereas, at low discriminability there was no change in latency, indicating that the contraction of the ISI had completely cancelled the effect of change in caution. Thus, the effects which would follow from variation of the ISI were strongest on response accuracy at high discriminability, and on response latency at low

discriminability. These are consistent with the relationship between discriminability level and the covariation which occurs in accuracy and sampling rate as the ISI is varied.

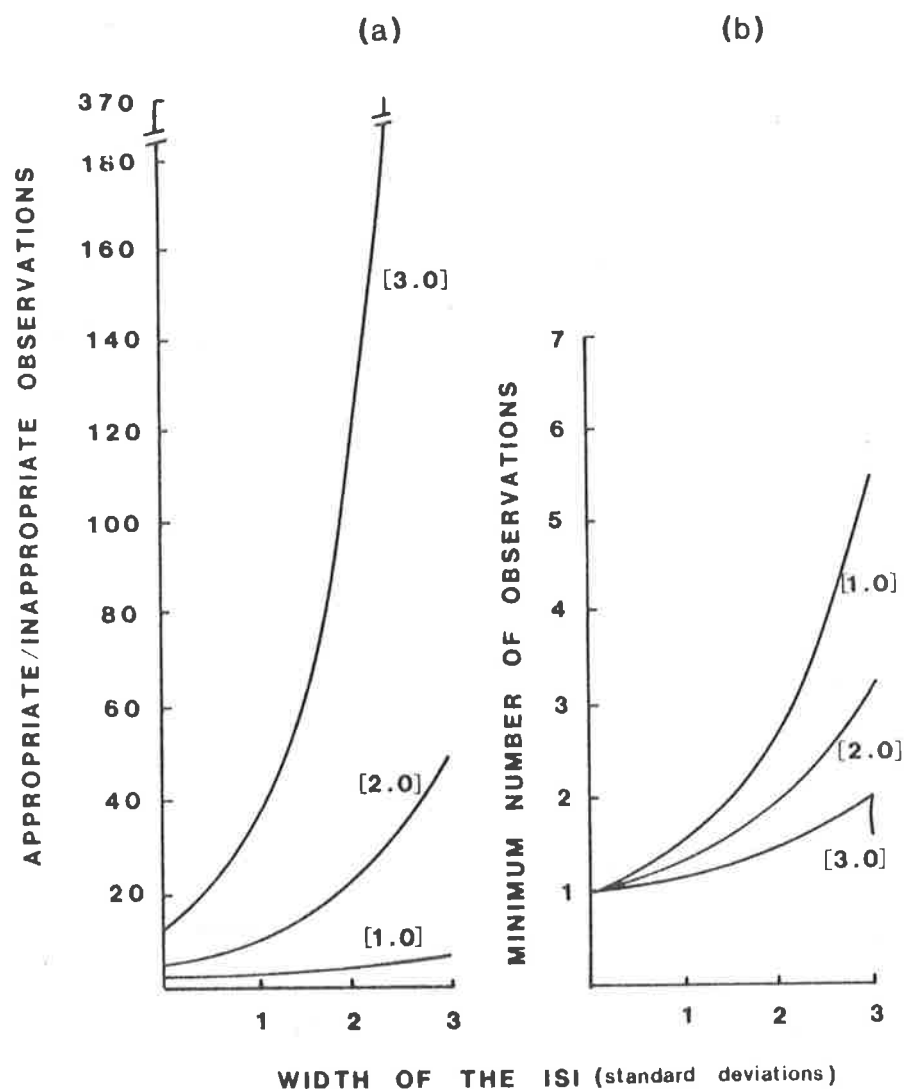


FIGURE 88

Relationship between the width of the interval of sensory indifference, ISI, and (a) the ratio of the proportion of observations which provide evidence for an appropriate response to the proportion which provide evidence for an inappropriate response, and (b) the expected number of observations until one is sampled outside the ISI. The ISI is located symmetrically about the intersection of sensory distributions whose means are separated by either 1.0, 2.0, or 3.0 standard deviations (given in square brackets by each curve).

(c) Previous results and the principle of 'least variation'

In the present experiment, the effects of sampling regulation did not emerge on response accuracy until there had been two step changes in discriminability. The fact that any one of three different sampling rates could occur at any time, one for each line length difference at each level of discriminability, may have made it exceedingly difficult for the regulation process to establish those sensory boundaries which would produce the desired rate of information flow. Specifically, even though orderly expansions and contractions of the ISI would have led to a constant average rate being maintained across different levels of discriminability, there was no way in which a common rate could be produced between trials at each level of discriminability. Attempts to produce such a common rate may have resulted in somewhat erratic adjustments over the early part of the task, since they could have included shifts in the location of the ISI as well as variation of its width. The latter adjustment would have ultimately been successful in altering the average rate of sampling in the required fashion, while the former would have led to even greater discrepancies in trial-to-trial variation in sampling rate so that such adjustments were abandoned. Varying the width of the ISI also seems the most likely manipulation to appear in a situation where the range of sensory intensities is expanding and contracting, as in the present experiment, and merely has to be made to correspond to the contractions and expansions that are experienced in this range. Also the mean of all sensory intensities, which was located at the intersection of the sensory distributions, did not vary, and such variation might usually be required for shifts to take place in the location of the ISI. The previous experiments produced changes in the mean as well as range of sensory intensity and, therefore, can serve to indicate the sort of variation that

can be expected in the ISI when these sensory characteristics vary in a way that affects sampling rate.

(i) Change in signal detectability. In a detection task with signals and nonsignals that are equally likely to occur, and underlying sensory distributions that are normal and of equal variance, the mean of all sensory intensities is located at the intersection of the two sensory distributions. In contrast to the present experiment, one distribution, that of noise, usually remains at a fixed location along the sensory continuum so that the position of the intersection of the two distributions will vary as signal detectability is varied. Hence, in Experiment 5 the decline in signal detectability produced a decline in mean sensory intensity as well as the range of sensory intensity. If the only goal of the sensory regulation process is to maintain a constant average rate of information flow, in this situation the ISI will contract as the range of sensory intensity contracts, but the position of the ISI will not be altered. If, however, the goal of sensory regulation also includes the minimization of trial-to-trial variation in sampling rate, then, to be successful, the ISI will have to be shifted so that it is always located around the intersection of the two sensory distributions. This is illustrated in Figure 89. When the signal distribution declines from s_1 to s_2 , the position of the ISI can be maintained at ISI_1 (Figure 89(b)), and there be contracted so that the combined areas of n and s_2 that are contained within the ISI will be the same as the previous combined areas of n and s_1 . The average sampling rate that is associated with n and s_2 will then be the same as that previously produced by n and s_1 . However, the area of s_2 contained within the ISI is greater than that of n , so that on signal trials there will be a slower rate of sampling observations outside the ISI than on nonsignal trials. This can only be remedied by shifting the ISI to

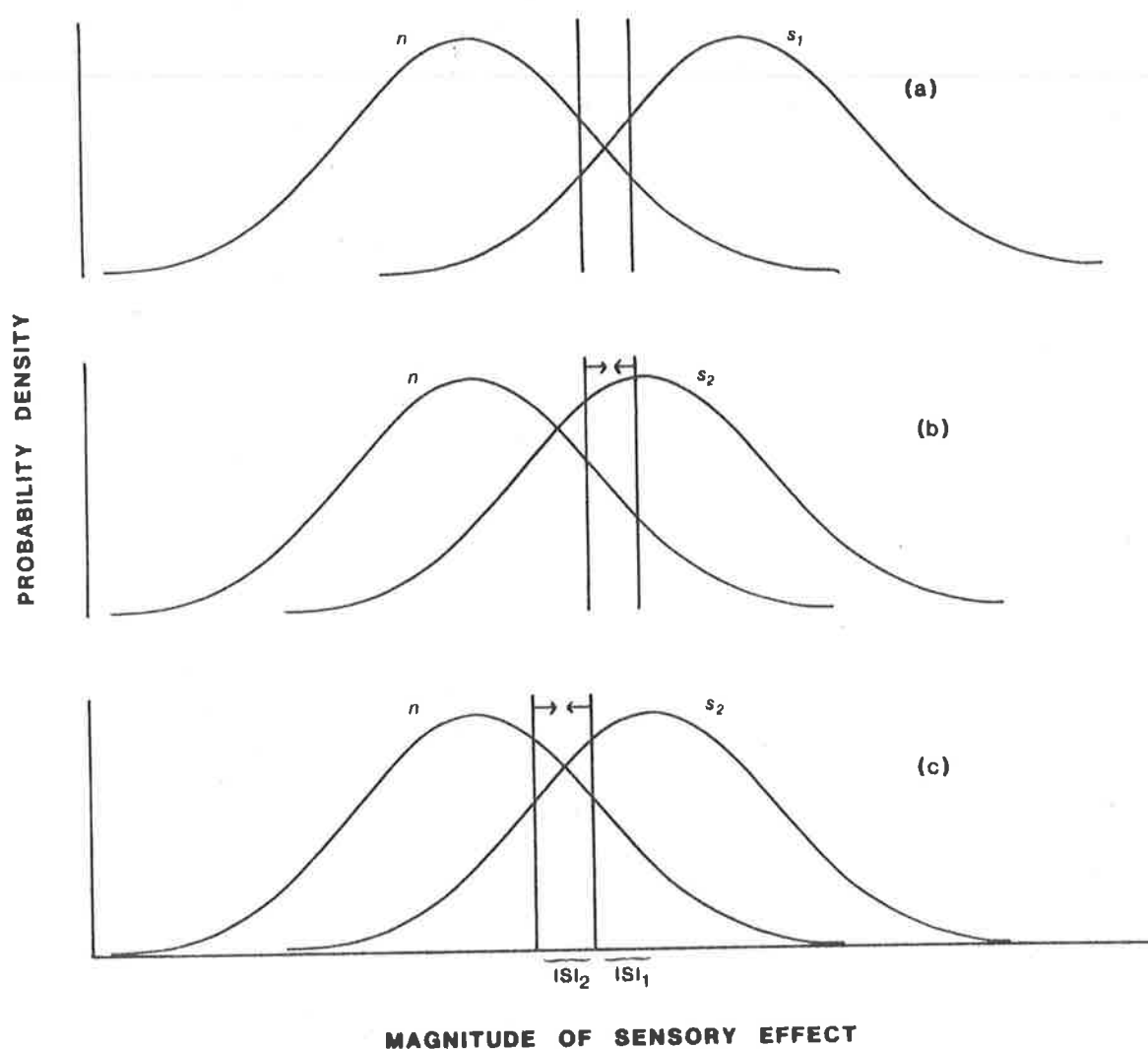


FIGURE 89 (a) Hypothetical distributions of the sensory effect produced by signals, s_1 , and nonsignals, n , and a symmetrically located interval of sensory indifference, ISI_1 . In (b) signal detectability has declined so that s is distributed as s_2 , and the location of the ISI remains at ISI_1 . In (c) the ISI remains at the intersection of the sensory distributions as detectability declines, and so is located at ISI_2 . The arrows in (b) and (c) indicate the shifts which must occur to the sensory boundaries in order to make the combined area of s and n within the ISI the same as that in (a).

ISI₂ so that equal areas of s_2 and n are located within the ISI (Figure 89(c)). Trial-to-trial variation in sampling rate will be minimized, while a simultaneous contraction of the ISI will cause the sampling rate to be maintained at the level which had been associated with n and s_1 .

If the ISI were to remain in a fixed location as the signal distribution moved downward then, as shown in Figure 89(b), progressively more of the signal distribution will become displaced to the left of the ISI. This will increase greatly the evidence in favour of nonsignal responses on signal trials and thereby decrease the probability and increase the latency of detections. The latency of detections will be further increased as a greater proportion of the signal distribution becomes located within the ISI. At the same time, decisions on nonsignal trials will only be affected by the contraction of the ISI that is necessary to offset the decreased sampling rate on signal trials. This will decrease the accuracy of nonsignal responses, but not to the extent that signal response accuracy is affected. Further, the decline in the probability of correct rejections, due to the contraction, will be accompanied by a decrease in their latency. These differences will be further heightened if there is intervention by the control accumulator. For instance, the increased evidence in favour of incorrect nonsignal responses, coupled with their increased probability, will increase the average confidence of nonsignal responses above its target level so that downward adjustments of the nonsignal criterion will be required. The decreasing evidence in favour of signal responses will at the same time increase the likelihood of upward adjustments to the signal response criterion. In this way, there will be an enhancement of the tendencies for nonsignal responses to become more probable and faster as signal responses become less probable and slower.

An ISI that shifts so that it remains around the intersection of the two sensory distributions, as shown in Figure 89(c), will produce changes that are similar on signal and nonsignal trials. The contraction of the ISI will maintain the rate at which sensory evidence enters the decision process, but the amounts that are sampled will become more comparable for the correct and incorrect response alternative on each trial. The accuracy and latency of both signal and nonsignal responses will then decline. Any intervention of the control accumulator will likewise affect signal and nonsignal responses in the same way, and performance trends will be similar to a decline in stimulus discriminability. The results of Experiment 5 confirm that both types of decision are affected in the same manner during a decline in signal detectability, except for the initial period when no change was evident on nonsignal trials and no variation of the ISI indicated. The later trends then suggest that the ISI was not only contracted as the range of sensory intensity contracted, but that its location was shifted to coincide with the mean sensory intensity within that range. The twofold effect of adjusting both the width and location of the ISI obeys a principle of 'least variation': the maintenance of a sampling rate so that it will vary minimally between trials.

(ii) Change in signal probability. Prior to the ramp reduction in signal probability during Experiments 1 to 4, signals and nonsignals were equally likely, so that the ISI again would have been located at the intersection of the signal and nonsignal distributions. The area of each distribution located within a symmetrically located ISI is not affected by change in the likelihood that the distribution will be sampled on any trial. Thus, sampling rates are not affected, so that, during the decline in signal probability, there would have been no reason for performance to change due

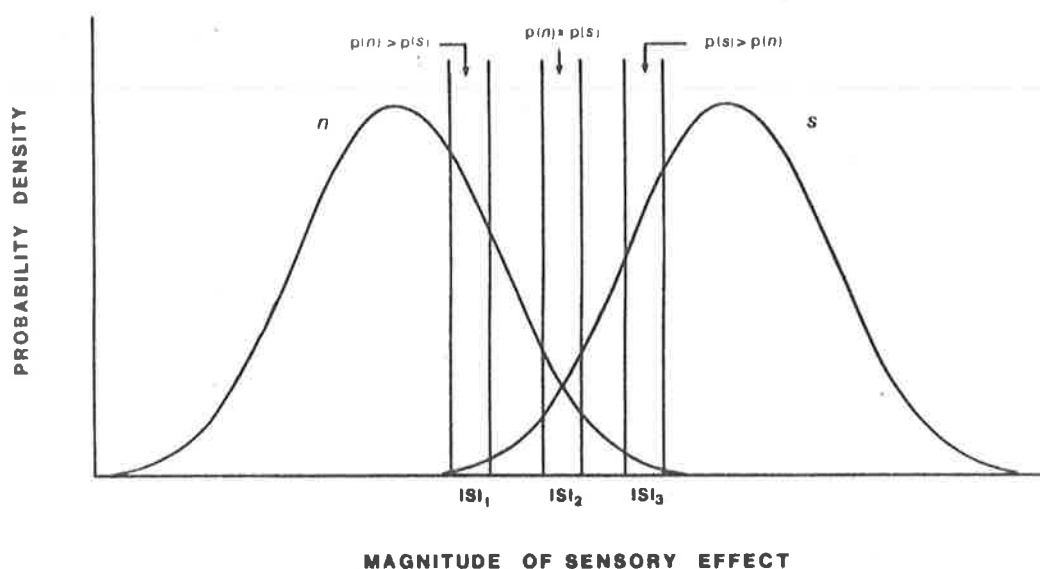


FIGURE 90 Hypothetical distributions of the sensory effect produced by signals, s , and nonsignals, n , and the location of the interval of sensory indifference, ISI, when the a priori probability of a signal is greater than, equal to, or less than that of a nonsignal, and the ISI is located around the mean magnitude of all observations.

to the regulation of the sampling process, provided the ISI retained its initial location. However, since regulation of the sampling process is considered to be independent of bias and caution in the decision process, and is only influenced by the characteristics of the sensory distributions from which evidence is sampled, there is appeal in a model in which the mean of all sensory intensities is employed as the boundary above which evidence accumulates towards one response alternative and below which it accumulates towards the other response alternative. That is, a range of values around mean sensory intensity are treated as having provided no evidence by which to determine the stimulus that is being sampled on any trial.

Such a tendency would account for the changes in performance observed during Experiments 1 to 4. As shown in Figure 90, if the ISI followed the mean intensity, then it would become displaced further to the left of the intersection of the signal and nonsignal distributions as

time progressed. Such a displacement will increase the ratio of observations favourable to a signal response on any trial, thereby increasing their probability and decreasing their latency. Conversely, nonsignal responses will decrease in probability and increase in latency. The increased evidence in favour of signal responses will increase the likelihood of downward adjustments of the signal criterion, and upward adjustments of the nonsignal criterion, in the control accumulator process. Such displacements of the ISI would also explain why Williges (1969) found that observers trained with an equal likelihood of signals and nonsignals adopted values of β greater than 1.0 when tested with a signal probability of .83, and less than 1.0 when tested with a signal probability of .17. In Figure 90, the conditional response probabilities that would be associated with ISI_3 , at the mean intensity for a signal probability greater than .5, yield a value of β greater than 1.0, while those associated with ISI_1 , at the mean intensity for a signal probability less than .5, yield a value of β less than 1.0.

A complication, of course, is that as an ISI becomes displaced away from the intersection of the sensory distributions, there will be a change in mean sampling rate as well as an increase in the trial-to-trial variation in sampling rate. For instance, as signal probability declines, if the ISI makes a shift comparable to that from ISI_2 to ISI_1 in Figure 90, there will be a decrease in average sampling rate as a greater portion of the nonsignal distribution produces empty observations, and nonsignal trials are now the most probable. At the same time, the smaller portion of the signal distribution now within the ISI will increase the sampling rate on signal trials, so that the variation between trials will increase. In order to increase the mean sampling rate, and also obey the principle of least variation, the ISI will have to be shifted back to its original location at the intersection of the sensory distributions rather than be contracted at its present

location. This re-adjustment of the ISI would tend to cancel the effects of the prior adjustment.

The ISI may not have been adjusted back towards its symmetrical position during Experiments 1 to 4, since the decline in signal probability would have continually moved mean sensory intensity further to the left on the sensory continuum. Even in Experiments 3 and 4, cumulatively-based signal probability, and therefore mean sensory intensity calculated over all preceding trials, continued to decrease during the period of stable signal probability which followed the ramp reduction. In Experiment 4, for example, cumulatively-based signal probability declined from .36 to .18 over the final thirteen trial blocks even though local signal probability was here held constant at .06. Similarly, in Williges' (1969) study, observers were given either 42 or 72 trials with an equal likelihood of signals and nonsignals before they were tested at a signal probability of .17 or .83. Thus, over the four main trial blocks of the 360 trial session, cumulatively-based signal probability declined from either .27 to .20 or from .31 to .22 at the low signal probability, and increased from either .73 to .80 or from .69 to .78 at the high signal probability. Hence, in these experiments the tendency for the ISI to be placed around mean sensory intensity, and shift with that intensity, would have conflicted with the adjustments necessary to minimize variation in sampling rate. However, experiments which hold signal probability constant at a value other than .5 will not create such a conflict and now will be considered in order to evaluate a model in which the ISI initially is placed at the mean sensory intensity and is subsequently regulated so that the trial-to-trial variability in sampling rate is minimized.

(d) Vigilance performance as affected by ISI location

According to the preceding model, the ISI will be located at a position other than the intersection of the sensory distributions when signal probability differs from .5. If this probability remains constant, any subsequent shifts in the ISI would indicate an attempt to reduce the variation in sampling rate between trials. The effects on performance of such re-locations of the ISI are unlikely to be found in tasks where observers have been extensively trained at the signal probability at which they are tested, since the sort of sensory adaptation that is implied by the process would have been completed. The effects will also be undetected if no time analysis is made of performance, irrespective of the amount of training. Thus, the effects cannot be expected in detection studies since they are usually interested in comparisons of mean performance between conditions and often employ a number of training sessions in order to stabilize performance to the greatest possible extent. However, in vigilance tasks, observers are usually trained and tested during the same experimental session, and the focus of interest is largely on the changes in performance which occur with time. Thus, it is in vigilance that the effects of sensory regulation, in the form of those early adjustments necessary to adapt to the task environment, are most likely to be observed.

(i) Vigilance increment versus decrement. When signal probability is held constant, the ISI initially will be located at mean sensory intensity, and then will be re-located toward the intersection of the two sensory distributions, in order to minimize trial-to-trial variation in sampling rate. As a consequence, the level of signal probability will determine whether a decrement, increment, or no change will be predicted for the detection parameter, β , during a vigilance task. For instance, when signal probability

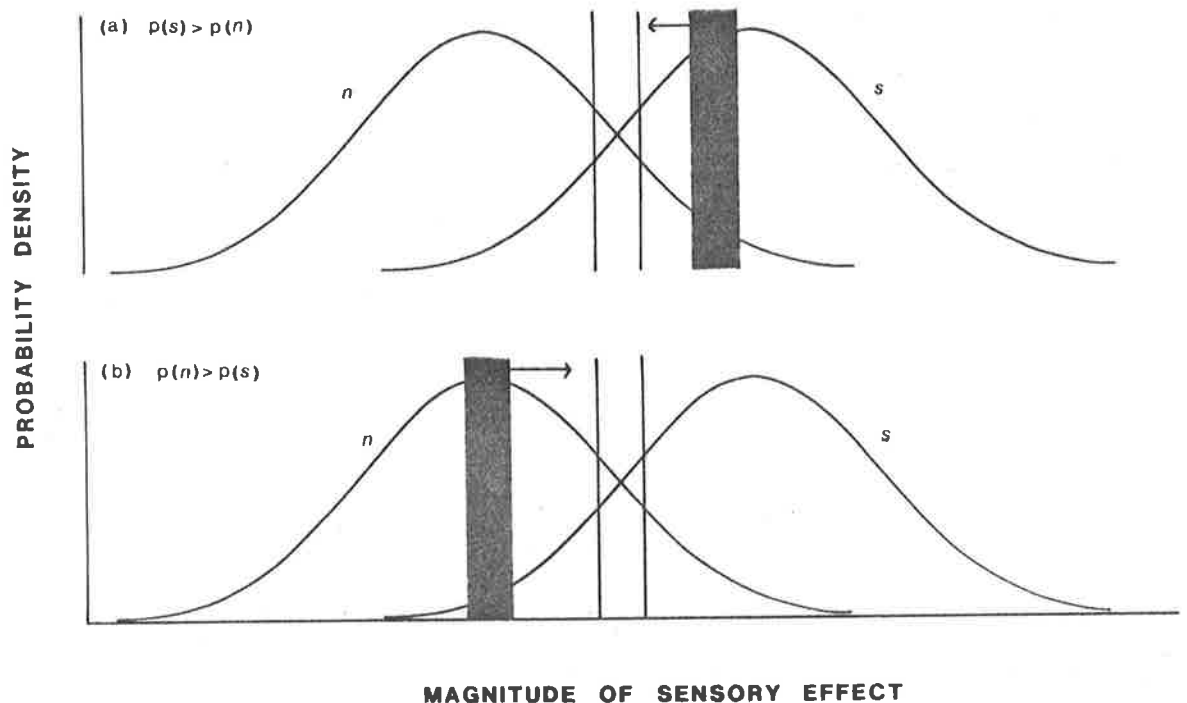


FIGURE 91 The initial position of an ISI (shaded) that is located around the value of mean sensory intensity and the direction in which it must be shifted, indicated by arrows, in order to reduce the difference in sampling rate between signal and nonsignal trials. In (a) signals are more likely than nonsignals and in (b) nonsignals are more likely than signals.

is greater than .5 the ISI will first be located to the right of the intersection and then be adjusted leftward, as shown in Figure 91(a). Such a shift will increase the evidence in favour of a signal response relative to a nonsignal response on both signal and nonsignal trials. If response bias remains constant, therefore, there will be an increase in the probability of both correct detections and false alarms, and this will be reflected as a decrease in β . Conversely, as shown in Figure 91(b), the balance of evidence in favour of a signal response will decrease when signal probability is less than .5, so that the conditional probabilities of detections and false alarms will decrease and β will increase. However, when signal probability

is .5 the ISI will always be located at the intersection of the sensory distributions, and β will be unaffected by the regulation of the sampling process. Thus, Williges (1969) found that, while β increased among observers who were trained and tested at a signal probability of .17 (the change in β usually associated with "vigilance decrement"), β decreased among observers trained and tested at a signal probability of .83. In two later studies, Williges (1971, 1973) again found increases in β during a vigilance task in which observers were trained and tested at a signal probability of .1, but no change in β when observers were trained and tested at a signal probability of .5.

Baddeley and Colquhoun (1969) found tendencies for β to increase among observers who had been trained and tested with signal probabilities that ranged from .02 to .36. The signal probability of .36 in that study is much higher than the values usually associated with performance decrements in vigilance, the infrequency of signals often being considered a defining characteristic of a vigilance task. However, it is still considerably less than .5, so that a tendency for β to increase would be predicted by the present model. It should also be noted that the optimal value of β for a signal probability of .36 is 1.78. Remarkably, the value observed by Baddeley and Colquhoun (1969) during training was 1.8. However, the value obtained during the first period of testing was 3.3, so that early change in β in no way reflected a tendency toward an optimal signal detection value. On the other hand, according to the present model, as the increase in β reflected a shift of the ISI toward its symmetrical location, it indicated a tendency toward more 'optimal' sampling behaviour. Only by placing the ISI at the intersection of the two sensory distributions can trial-to-trial variability in sampling rate be effectively minimized while the mean sampling

rate is simultaneously regulated. Thus, by satisfying the principle of least variation, the above trends in β indicate more ideal sampling.

(ii) Concurrent change in bias. When signal probability differs from .5, the initial location of the ISI is on the opposite side of the intersection to the optimal position of a signal detection decision criterion. As a result, the conditional probabilities of sampling evidence in favour of a signal response could be deemed "inappropriate" to the actual signal probability. However, the target levels of confidence that are set for responses, and any ensuing bias in decision criteria, will lead the observer to produce response rates appropriate to the signal probability employed.

As well as affecting sampling regulation, any change in signal probability can also influence performance by the effect it has on response confidence, and the potential thereof for an alteration in criterion levels. An example of the effects of change in bias occurring in conjunction with those of change in the location of the ISI is given by the study of Colquhoun and Baddeley (1967). One group of observers was trained at a signal probability of .02 and tested at one of .18, while another group was trained at .18 and tested at .02. Both test probabilities are quite low, so that a shift of the ISI toward the right on the sensory continuum might be expected, which would tend to increase β . However, as probability changed by a factor of nine from training to testing, a sizeable change was produced in the ratio of detections to false alarms, that is, signal responses made at higher levels of confidence relative to those made at lower levels of confidence. An increase in signal probability would increase this ratio, so that the average confidence of signal responses would exceed the level established in training. Thus a decrease in the criterion for a signal response would be required. Conversely, a decrease in signal probability would decrease

the proportion of signal responses made at higher levels of confidence, and so require an increase in the signal criterion in order to restore signal response confidence to its target level. The decrease in the criterion at the higher signal probability would counter the effects of ISI re-location, while the increase at low signal probability would enhance the effects. This agrees with the decrease in β which Colquhoun and Baddeley (1967) found when probability shifted from .02 to .18, and the increase found when probability shifted from .18 to .02.

Another interesting feature of Colquhoun and Baddeley's (1967) study was a decrement over time in detection rate among observers trained and tested at the signal probability of .18, but not among those trained and tested at the .02 probability. This seems ironic, since the re-location of the ISI should be more drastic when mean intensity is furthest from the intersection, that is, at the more extreme, and in this case, lower, signal probability. However, observers had been given 200 training trials at the .02 signal probability compared with 100 at the .18 probability. Hence, if re-location of the ISI is an adaptation process which occurs fairly early in training and testing, it could have been completed by the time the observers at the lower signal probability encountered the experimental trials.

(e) Vigilance performance as affected by ISI width

The present discussion must be deemed highly speculative, especially in the present state of development of the model. A full mathematical elaboration, combined with simulations of the hypothetical process, would be required in order to test fully the proposals which have been presented. Before leaving the topic of vigilance, however, we may perhaps speculate on one final aspect - the influence of the sampling process on the detection parameter d' . The previous section attempted to demonstrate the manner

in which the detection measure of bias, β , could be affected by shifts in the location of the ISI. The other characteristic of the ISI, its width, will affect the detection measure of sensitivity, d' , and, assuming a regulation process which seeks to maintain the rate at which observations are sampled outside the ISI, the general level of performance during a vigilance task.

(i) Duration of events. In section (b), above, it was shown that a contraction of the ISI, in order to increase sampling rate, would lower the accuracy of performance, so that d' will be a direct function of the width of the ISI. Thus, those factors which determine the rate at which information must be sampled in a particular task, and so prescribe the necessary width of the ISI, will directly affect the observer's ability to detect signals at any given level of signal intensity. One obvious factor is the duration of each stimulus presentation. At very low durations, the ISI will have to be narrower than at longer durations, simply to ensure that some information is passed on to the decision process during each stimulus presentation. Thus, for a given signal intensity and response criteria levels, d' can be expected to be higher at longer signal durations. Davies and Tune (1970, Chapter 3), cite a number of studies which show that detection performance is higher at longer signal durations. Further, J. Mackworth and Taylor (1963) found that an increase in d' was associated with the increased ability to detect signals of .38 seconds durations relative to signals of the same type with a duration of .32 seconds. Also, in a task in which event rate and event duration were confounded, J. Mackworth (1965) reported initially higher values of d' with event rates of 200 per minute ($d' = 3.0$) as opposed to 40 per minute ($d' = 2.2$). However, the higher event rate had an event duration of .15 seconds, while the duration was only .08 seconds

at the low event rate. The cost of any observation yielding no evidence toward a decision would have been much greater in the latter case, where only one observation may have occurred during each stimulus presentation.

(ii) Event rate. Unlike the last study cited, in which event duration was purposely varied with event rate, most studies have found superior performance to be associated with lower event rates (e.g. Jerison, Pickett & Stenson, 1965; Loeb & Binford, 1968; Taub & Osborne, 1968; Colquhoun, 1969; Parasuraman & Davies, 1976). For a given event duration, sampling behaviour should not be affected by the rate at which discrete events occur - provided, of course, that responses do not occasionally interfere with the sampling process at higher event rates, giving rise to a need to generally increase the rate at which information is sampled. In this case, decision criteria could be lowered, or the ISI contracted. Either of these adjustments would result in faster decisions, and a lower likelihood that any response which is triggered would interfere with the next stimulus presentation. Since the observer would be aware of any 'conflict between seeing and responding', its resolution would most likely occur through a lowering of criteria.

However, there might also be instances where the ISI has to be contracted at higher event rates. If the full stimulus presentation is analyzed prior to a decision, then, since there are very small intervals between stimuli at very high event rates, there might be an overlap between the process of passing on information regarding a previous stimulus (or even the later clearing of accumulators) and the presentation of the next stimulus. The possibility of this sort of overlap at short inter-stimulus intervals seems great when consideration is given to the sampling of a neural trace which occurs following stimulus termination (Sperling, 1960, 1963), and which

further occupies the interstimulus interval. By increasing the sampling rate, a contraction of the ISI would allow a decision to be based on less sensory evidence than is contained in a full stimulus presentation. A decision could therefore be based on evidence gathered up until some finite moment prior to stimulus termination. In this way, a stimulus could be effectively 'clipped' prior to its termination, which would not be necessary with longer intervals between stimuli.

Contractions of the ISI at higher event rates might have occurred in tasks like those of Taub and Osborne (1968), in which observers were required to monitor discrete jumps of the second hand of a clock in order to detect occasional jumps which made a longer sweep across the clock face. At the highest event rate in the task, two stimuli were presented during each second of time, compared with one stimulus every four seconds at the lower rate. Thus, a clash between processing one jump and monitoring the next one is likely to have occurred at the higher event rate, unless the observer could contract the ISI (each 'observation' may have been of angular velocity), and thereby decide that a jump had exceeded a critical length prior to the actual completion of the jump. At the lower rates, however, a full jump could be observed, and a decision reached following this.

C. SOME GENERAL CONCLUSIONS

1. The Nature of Vigilance Decrement

The notion that vigilance decrement can be attributed, at least in part, to an adaptation process at the sensory level, seems likely to be somewhat controversial, as does the idea that this process occurs during the earliest stages of task experience. However, vigilance decrement as a phenomenon of the early stages of task performance is in no way a recent concept. Two decades have passed since Jerison (1963) presented evidence that the vigilance decrement is much more rapid than had been assumed by most theorists at that time. Further, as was suggested by Jerison, the vigilance decrement may not be as ubiquitous as is often thought. It could even be the case that a vigilance decrement will fail to occur, since it is often as contrived as it is real. If the original inter-signal schedule of N. Mackworth (1950) is examined, it becomes readily apparent that the signal rate during the first five minutes of the task is much higher than the rate for the task as a whole. Colquhoun and Baddeley (1964, 1967) have aptly demonstrated the consequences of such false expectancies early in the task on the later behaviour of the observer. One is even led to contemplate the possibility that the reason the "Mackworth schedule" gained such widespread popularity in the earliest days of vigilance research rested in the fact that it produced a vigilance decrement as much as in the fact of its convenience. The present-day use of computers to generate experimental stimuli, and the degree of randomization that this makes possible, has probably contributed to the subtle shift of interest in vigilance research away from those factors which influence the decline in attentiveness during a vigil to those factors which influence the level of attentiveness.

2. The Proposed Model of Performance

The present model is by no means a simple one. However, a more rigorous mathematical treatment than was attempted here would allow it to be properly evaluated, and such an evaluation would also allow it to achieve a level of predictive ability which may prove to be worth such an exercise. This follows, since it should provide the means by which to measure the predicted response by each mechanism to variation in specific factors, such as stimulus probability, stimulus discriminability, and, perhaps, the general level of sensory excitation. Further, it will establish the ranges within which these responses can be expected to produce summative or interactive effects on any given response measure.

The parallel-process model differs in an important way from all of the previous models considered in this thesis. Rather than a hierarchy of regulatory processes, in which control is "passed" between processes in such a fashion that they function sequentially, the present model proposes that control may occur simultaneously at different levels of information processing. Also, it had previously been assumed that different reference variables could be responsible for the variation shown at different times in a single parameter of the decision process. For example, adjustments of a decision criterion were thought to occur because of change in either response probability or response confidence. In contrast, the present model specifies a unique characteristic of the decision process to be altered by each regulatory process. Further, a directionality in the effects of control action is implied by the model: while the regulation of sensory sampling and the action of the control accumulator represent parallel and independent processes, sampling regulation can produce changes in performance that may require subsequent action to be taken by the higher accumulator process.

However, the adjustment of decision criteria will not affect the rate at which information enters the decision process. That is, the lower system in the information process can indirectly influence the behaviour of the higher system, but the actions of the higher system will not produce changes that extend as far as the level of the lower system.

Lastly, although plausible arguments can be raised to question the necessity of specifying a regulatory system at the sensory level, there may be an evolutionary necessity for the development and control of the process which it regulates. An "interval of sensory indifference" essentially implies that certain sensory events, those most common to all stimuli, are prohibited from entering that process by which stimuli are differentiated. As such, sampling regulation may indeed be compared to a self-governing filtering process, which restricts the entry of redundant sensory information to higher-order analytic processes.

REFERENCES

- ADAMS, J.A. Vigilance in the detection of low-intensity visual stimuli. *Journal of Experimental Psychology*, 1956, 52, 204-208.
- ADAMS, J.A., & BOULTER, L.R. An evaluation of the activationist hypothesis of vigilance. *Journal of Experimental Psychology*, 1962, 64, 495-504.
- ADAMS, J.A., & BOULTER, L.R. Spatial and temporal uncertainty as determinants of vigilance behavior. *Journal of Experimental Psychology*, 1964, 67, 127-131.
- ATTNEAVE, F. Psychological probability as a function of perceived frequency. *Journal of Experimental Psychology*, 1953, 46, 81-86.
- AUDLEY, R.J. A stochastic model of individual choice behaviour. *Psychological Review*, 1960, 67, 1-15.
- BADDELEY, A.D., & COLQUHOUN, W.P. Signal probability and vigilance: A reappraisal of the 'signal-rate' effect. *British Journal of Psychology*, 1969, 60, 169-178.
- BAKAN, P. Discrimination decrement as a function of time in a prolonged vigil. *Journal of Experimental Psychology*, 1955, 50, 387-390.
- BAKAN, P. Extraversion-introversion and improvement in an auditory vigilance task. *British Journal of Psychology*, 1959, 50, 325-332.
- BAKAN, P. Effect of meprobamate on auditory vigilance. *Perceptual and Motor Skills*, 1961, 12, 26.
- 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. 22-28.
- BAKER, C.H. Towards a theory of vigilance. *Canadian Journal of Psychology*, 1959, 13, 35-42.
- BAKER, C.H. Maintaining the level of vigilance by means of artificial signals. *Journal of Applied Psychology*, 1960, 44, 336-338. (a)
- BAKER, C.H. Observing behavior in a vigilance task. *Science*, 1960, 132, 674-675. (b)
- BAKER, C.H. Maintaining the level of vigilance by means of knowledge of results about a secondary task. *Ergonomics*, 1961, 4, 311-316.
- BAKER, C.H. Further toward a theory of vigilance. In D.N. Buckner and J.J. McGrath (Eds.), *Vigilance: A symposium*. New York: McGraw-Hill, 1963, pp. 127-154.

- BEACH, L.R. A note on the intrasubject similarity of subjective probabilities obtained by estimates and by bets. *Organizational Behavior and Human Performance*, 1974, 11, 250-252.
- BEACH, L.R., ROSE, R.M., SAYEKI, Y., WISE, J.A., & CARTER, W.B. Probability learning: Response proportions and verbal estimates. *Journal of Experimental Psychology*, 1970, 86, 165-170.
- BEACH, L.R., & WISE, J.A. Subjective probability revision and subsequent decisions. *Journal of Experimental Psychology*, 1969, 81, 561-565.
- BERGUM, B.O., & LEHR, D.J. End-spurt in vigilance. *Journal of Experimental Psychology*, 1963, 66, 383-385.
- BINFORD, J.R., & LOEB, M. Changes within and over repeated sessions in criterion and effective sensitivity in an auditory vigilance task. *Journal of Experimental Psychology*, 1966, 72, 339-345.
- BLAKE, M.J.F. Time of day effects on performance in a range of tasks. *Psychonomic Science*, 1967, 9, 349-350.
- BOWEN, H.M. Vigilance as a function of signal frequency and flash rate. *Perceptual and Motor Skills*, 1964, 18, 333-338.
- BROADBENT, D.E. Classical conditioning and human watch-keeping. *Psychological Review*, 1953, 60, 331-339.
- BROADBENT, D.E. *Decision and stress*. London: Academic Press, 1971.
- BROADBENT, D.E. Levels, hierarchies and the locus of control. *Quarterly Journal of Experimental Psychology*, 1977, 29, 181-201.
- BROADBENT, D.E., & GREGORY, M. Vigilance considered as a statistical decision. *British Journal of Psychology*, 1963, 54, 309-323.
- BROADBENT, D.E., & GREGORY, M. Effects of noise and of signal rate upon vigilance analyzed by means of decision theory. *Human Factors*, 1965, 7, 155-162.
- BUCK, L. Reaction time as a measure of perceptual vigilance. *Psychological Bulletin*, 1966, 65, 291-304.
- CARTERETTE, E.C., FRIEDMAN, M.P., & COSMIDES, R. Reaction-time distributions in the detection of weak signals in noise. *Journal of the Acoustical Society of America*, 1965, 38, 531-542.
- CATALANO, J.F. Effect of perceived proximity to end of task upon end-spurt. *Perceptual and Motor Skills*, 1973, 36, 363-372.
- COLQUHOUN, W.P. The effect of a short rest pause on inspection efficiency. *Ergonomics*, 1959, 2, 367-372.

- COLQUHOUN, W.P. The effect of "unwanted" signals on performance in a vigilance task. *Ergonomics*, 1961, 4, 41-51.
- COLQUHOUN, W.P. The effect of 'unwanted' signals on performance in a vigilance task: A reply to Jerison. *Ergonomics*, 1966, 9, 417-419.
- COLQUHOUN, W.P. Effects of raised ambient temperature and event rate on vigilance performance. *Aerospace Medicine*, 1969, 40, 413-417.
- COLQUHOUN, W.P., & BADDELEY, A.D. Role of pretest expectancy in vigilance decrement. *Journal of Experimental Psychology*, 1964, 68, 156-160.
- COLQUHOUN, W.P., & BADDELEY, A.D. Influence of signal probability during pretraining on vigilance decrement. *Journal of Experimental Psychology*, 1967, 73, 153-155.
- COLQUHOUN, W.P., & EDWARDS, R.S. Practice effects on a visual vigilance task with and without search. *Human Factors*, 1970, 12, 537-545.
- COLQUHOUN, W.P., & GOLDMAN, R.F. Vigilance under induced hyperthermia. *Ergonomics*, 1972, 15, 621-632.
- CORCORAN, D.W.J. Individual differences in performance after loss of sleep. Ph.D. Thesis (unpublished), University of Cambridge, 1962. (a)
- CORCORAN, D.W.J. Noise and loss of sleep. *Quarterly Journal of Experimental Psychology*, 1962, 14, 178-182. (b)
- CORCORAN, D.W.J. Doubling the rate of signal presentation in a vigilance task during sleep deprivation. *Journal of Applied Psychology*, 1963, 47, 412-415.
- CORCORAN, D.W.J. Changes in heart rate and performance as a result of loss of sleep. *British Journal of Psychology*, 1964, 55, 307-314. (a)
- CORCORAN, D.W.J. The relation between introversion and salivation. *American Journal of Psychology*, 1964, 77, 298-300. (b)
- CORCORAN, D.W.J. Personality and the inverted-U relation. *British Journal of Psychology*, 1965, 56, 267-273.
- CRAIG, A. Is the vigilance decrement simply a response adjustment towards probability matching? *Human Factors*, 1978, 20, 441-446.
- CRAIG, A. Decision processes in detection and recognition: Models and influencing variables. In M.A. Sinclair and J.N. Clare (Eds.), *Search and the human observer*. London: Taylor and Francis, 1979, pp. 4-13.

- CURRY, R.E., NAGEL, D.C., & GAI, E.G. Decision behavior with changing signal strength. *Journal of Mathematical Psychology*, 1977, 15, 46-69.
- DAVIES, D.R., & HOCKEY, G.R.J. The effects of noise and doubling the signal frequency on individual differences in visual vigilance performance. *British Journal of Psychology*, 1966, 57, 381-389.
- DAVIES, D.R., HOCKEY, G.R.J., & TAYLOR, A. Varied auditory stimulation, temperament differences and vigilance performance. *British Journal of Psychology*, 1969, 60, 453-457.
- DAVIES, D.R., LANG, L., & SHACKLETON, V.J. The effects of music and task difficulty on performance at a visual vigilance task. *British Journal of Psychology*, 1973, 64, 383-389.
- DAVIES, D.R., & TUNE, G.S. *Human vigilance performance*. London: Staples Press, 1970.
- DEESE, J. Some problems in the theory of vigilance. *Psychological Review*, 1955, 62, 359-368.
- DUCHARME, W.M. Response bias explanation of conservative human inference. *Journal of Experimental Psychology*, 1970, 85, 66-74.
- DUCHARME, W.M., & DONNELL, M.L. Intrасubject comparison of four response modes for "subjective probability" assessment. *Organizational Behavior and Human Performance*, 1973, 10, 108-117.
- EDWARDS, W. Probability learning in 1000 trials. *Journal of Experimental Psychology*, 1961, 62, 385-394.
- EGAN, J.P., GREENBERG, G.Z., & SCHULMAN, A.I. Operating characteristics, signal detectability, and the method of free response. *Journal of the Acoustical Society of America*, 1961, 33, 993-1007.
- EMBREY, D.E. Training the inspector's sensitivity and response strategy. In C.G. Drury and J.G. Fox (Eds.), *Human reliability in quality control*. London: Taylor and Francis, 1975, pp. 123-131.
- EMMERICH, D.S., GRAY, J.L., WATSON, C.S., & TANIS, D.C. Response latency, confidence, and ROCs in auditory signal detection. *Perception and Psychophysics*, 1972, 11, 65-72.
- FREEMAN, P.R. *Tables of d' and β* . Medical Research Council Applied Psychology Research Unit Report, No. 529, November, 1964.
- GESCHIEDER, G.A., WRIGHT, J.H., & EVANS, M.B. Reaction time in the detection of vibrotactile signals. *Journal of Experimental Psychology*, 1968, 77, 501-504.

- GESCHEIDER, G.A., WRIGHT, J.H. & POLAK, J.W. Detection of vibrotactile signals differing in probability of occurrence. *The Journal of Psychology*, 1971, 78, 253-260.
- GESCHEIDER, G.A., WRIGHT, J.H., WEBER, B.J., & BARTON, W.G. Absolute thresholds in vibrotactile signal detection. *Perception and Psychophysics*, 1971, 10, 413-417.
- GESCHEIDER, G.A., WRIGHT, J.H., WEBER, B.J., KIRCHNER, B.M., & MILLIGAN, E.A. Reaction time as a function of the intensity and probability of occurrence of vibrotactile signals. *Perception and Psychophysics*, 1969, 5, 18-20.
- GREEN, D.M., & SWETS, J.A. *Signal detection theory and psychophysics*. New York: Wiley and Sons, 1966.
- GURALNICK, M.J. Observing responses and decision processes in vigilance. *Journal of Experimental Psychology*, 1972, 93, 239-244.
- GURALNICK, M.J. Effects of event rate and signal difficulty on observing responses and detection measures in vigilance. *Journal of Experimental Psychology*, 1973, 99, 261-265.
- GURALNICK, M.J., & HARVEY, K.G. Response requirements and performance in a visual vigilance task. *Psychonomic Science*, 1970, 20, 215-217.
- HAMILTON, P., WILKINSON, R.T., & EDWARDS, R.S. A study of four days partial sleep deprivation. In W.P. Colquhoun (Ed.), *Aspects of human efficiency*. London: English Universities Press, 1972, pp. 101-112.
- HEBB, D.O. Drives and the CNS (conceptual nervous system). *Psychological Review*, 1955, 62, 243-254.
- HELSON, H. Adaptation-level as frame of reference for prediction of psychophysical data. *American Journal of Psychology*, 1947, 60, 1-29.
- HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychological Review*, 1948, 55, 297-313.
- HOLLAND, J.G. Human vigilance. *Science*, 1958, 128, 61-67.
- HOLLAND, J.G. Human vigilance. In D.N. Buckner and J.J. McGrath (eds.), *Vigilance: A symposium*. New York: McGraw-Hill, 1963, pp. 247-264.
- HOWELL, W.C. Intuitive "counting" and "tagging" in memory. *Journal of Experimental Psychology*, 1970, 85, 210-215.
- HOWELL, W.C. Representation of frequency in memory. *Psychological Bulletin*, 1973, 80, 44-53.

- INDLIN, Yu.A. The observer as a system with feedback. In H.-G. Geissler and Yu.M. Zabrodin (Eds.), *Advances in psychophysics*. Berlin: VEB Deutscher Verlag der Wissenschaften, 1976, pp. 175-193.
- JENKINS, H.M. The effect of signal-rate on performance in visual monitoring. *The American Journal of Psychology*, 1958, 71, 647-661.
- JERISON, H.J. On the decrement function in human vigilance. In D.N. Buckner and J.J. McGrath (Eds.), *Vigilance: A symposium*. New York: McGraw-Hill, 1963, pp. 199-212.
- JERISON, H.J. Human and animal vigilance. *Perceptual and Motor Skills*, 1965, 21, 580-582.
- JERISON, H.J. Signal detection theory in the analysis of human vigilance. *Human Factors*, 1967, 9, 285-288.
- JERISON, H.J. Vigilance: A paradigm and some physiological speculations. *Acta Psychologica*, 1970, 33, 367-380.
- JERISON, H.J., PICKETT, R.M., & STENSON, H.H. The elicited observing rate and decision processes in vigilance. *Human Factors*, 1965, 7, 107-128.
- JOHNSON, E.M., & PAYNE, M.C. Jr. Vigilance: Effects of frequency of knowledge of results. *Journal of Applied Psychology*, 1966, 50, 33-34.
- JOHNSTON, W.A., HOWELL, W.C., & GOLDSTEIN, I.L. Human vigilance as a function of signal frequency and stimulus density. *Journal of Experimental Psychology*, 1966, 72, 736-743.
- JOHNSTON, W.A., HOWELL, W.C., & WILLIGES, R.C. The components of complex monitoring. *Organizational Behavior and Human Performance*, 1969, 4, 112-124.
- KAPPAUF, W.E., & POWE, W.E. Performance decrement at an audio-visual checking task. *Journal of Experimental Psychology*, 1959, 57, 49-56.
- KATZ, L. A comparison of type II operating characteristics derived from confidence ratings and from latencies. *Perception and Psychophysics*, 1970, 8, 65-68.
- KEISTER, M.E., & MCLAUGHLIN, R.J. Vigilance performance related to extraversion-introversion and caffeine. *Journal of Experimental Research in Personality*, 1972, 6, 5-11.
- KLEITMAN, N. *Sleep and wakefulness*. Chicago: University of Chicago Press, 1939.
- LA BERGE, D. A recruitment theory of simple behavior. *Psychometrika*, 1962, 27, 375-396.

- LEVINE, J.M. The effects of values and costs on the detection and identification of signals in auditory vigilance. *Human Factors*, 1966, 8, 525-537.
- LISPER, H.-O. Classical reaction time and signal rate in a vigilance setting. *Psychonomic Science*, 1969, 17, 217-218.
- LISPER, H.-O., & ERICSSON, S. Effects of signal frequency on increase in reaction time in a 10-minute auditory monitoring task. *Journal of Experimental Psychology*, 1973, 98, 316-319.
- LISPER, H.-O., KJELBERG, A., & MELIN, L. Effects of signal intensity on increase of reaction time on an auditory monitoring task. *Perceptual and Motor Skills*, 1972, 34, 439-444.
- LOEB, M., & BINFORD, J.R. Vigilance for auditory intensity changes as a function of preliminary feedback and confidence level. *Human Factors*, 1964, 6, 445-458.
- LOEB, M., & BINFORD, J.R. Variation in performance on auditory and visual monitoring as a function of signal and stimulus frequencies. *Perception and Psychophysics*, 1968, 4, 361-367.
- LOEB, M., HAWKES, G.R., EVANS, W.O., & ALLUISI, E.A. The influence of d-amphetamine, benactyzine, and chlorpromazine on performance in an auditory vigilance task. *Psychonomic Science*, 1965, 3, 29-30.
- LUCE, R.D., & GREEN, D.M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, 79, 14-57.
- MACKIE, R.R. (Ed.) *Vigilance: Theory, operational performance, and physiological correlates*. New York: Plenum Press, 1977.
- MACKWORTH, J.F. The effect of true and fake knowledge of results on the detectability of signals in a vigilance task. *Canadian Journal of Psychology*, 1964, 18, 209-223.
- MACKWORTH, J.F. Deterioration of signal detectability during a vigilance task as a function of background event rate. *Psychonomic Science*, 1965, 3, 421-422.
- MACKWORTH, J.F. The effect of signal rate on performance in two kinds of vigilance task. *Human Factors*, 1968, 10, 11-18.
- MACKWORTH, J.F. *Vigilance and habituation*. Victoria: Penguin Books, 1969.
- MACKWORTH, J.F. *Vigilance and attention*. Middlesex: Penguin Books, 1970.

- MACKWORTH, J.F., & TAYLOR, M.M. The d' measure of signal detectability in vigilance-like situations. *Canadian Journal of Psychology*, 1963, 17, 302-325.
- MACKWORTH, N.H. The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1948, 1, 6-21.
- MACKWORTH, N.H. *Researches on the measurement of human performance*. Medical Research Council Special Report Series, H.M. Stationary Office, London, No. 268, 1950.
- MACKWORTH, N.H., KAPLAN, I.T., & METLAY, W. Eye movements during vigilance. *Perceptual and Motor Skills*, 1964, 18, 397-402.
- MALMO, R.B. Activation: A neuropsychological dimension. *Psychological Review*, 1959, 66, 367-386.
- MARTZ, R.L. Auditory vigilance as affected by signal rate and intersignal interval variability. *Perceptual and Motor Skills*, 1967, 24, 195-203.
- MCFARLAND, D.J. *Feedback mechanisms in animal behaviour*. London: Academic Press, 1971.
- MILOŠEVIĆ, S. Effect of time and space uncertainty on a vigilance task. *Perception and Psychophysics*, 1974, 15, 331-334.
- MURRELL, G.A. A reappraisal of artificial signals as an aid to a visual monitoring task. *Ergonomics*, 1975, 18, 693-700.
- PARASURAMAN, R., & DAVIES, D.R. Decision theory analysis of response latencies in vigilance. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 578-590.
- PARDUCCI, A., & SANDUSKY, A.J. Limits on the applicability of signal detection theories. *Perception and Psychophysics*, 1970, 7, 63-64.
- PARKS, T.E. Signal-detectability theory of recognition-memory performance. *Psychological Review*, 1966, 73, 44-58.
- PEARSON, E.S., & HARTLEY, H.D. *Biometrika tables for statisticians*. Vol. 1. London: Cambridge University Press, 1966.
- PETERSON, C.R., & DUCHARME, W.M. A primacy effect in subjective probability revision. *Journal of Experimental Psychology*, 1967, 73, 61-65.
- PETERSON, C.R., & MILLER, A.J. Sensitivity of subjective probability revision. *Journal of Experimental Psychology*, 1965, 70, 117-121.

- PETERSON, C.R., SCHNEIDER, R.J., & MILLER, A.J. Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology*, 1965, 69, 522-527.
- PETERSON, C.R., ULEHLA, Z.F., MILLER, A.J., BOURNE, L.E., & STILSON, D.W. Internal consistency of subjective probabilities. *Journal of Experimental Psychology*, 1965, 70, 526-533.
- PHILLIPS, L.D., & EDWARDS, W. Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 1966, 72, 346-354.
- PIKE, R. Response latency models for signal detection. *Psychological Review*, 1973, 80, 53-68.
- PIKE, R., DALGLEISH, L. The components of latency of response in two models for auditory detection with deadlines. *Perception and Psychophysics*, 1976, 19, 231-239.
- PIKE, R., MCFARLAND, K., & DALGLEISH, L. Speed-accuracy tradeoff models for auditory detection with deadlines. *Acta Psychologica*, 1974, 38, 379-399.
- PIKE, R., & RYDER, P. Response latencies in the yes/no detection task: An assessment of two basic models. *Perception and Psychophysics*, 1973, 13, 224-232.
- PITZ, G.F., DOWNING, L., & REINHOLD, H. Sequential effects in the revision of subjective probabilities. *Canadian Journal of Psychology*, 1967, 21, 381-393.
- POWERS, W.T., CLARK, R.K., & MCFARLAND, R.L. A general feedback theory of human behavior: Part I. *Perceptual and Motor Skills*, 1960, 11, 71-88. (a)
- POWERS, W.T., CLARK, R.K., & MCFARLAND, R.L. A general feedback theory of human behavior: Part II. *Perceptual and Motor Skills*, 1960, 11, 309-323. (b)
- RABBIT, P. Current paradigms and models in human information processing. In V. Hamilton and D.M. Warburton (Eds.), *Human stress and cognition: An information processing approach*. New York: Wiley and Sons, 1979, pp. 115-140.
- ROBINSON, G.H. Continuous estimation of a time-varying probability. *Ergonomics*, 1964, 7, 7-21.
- SMITH, R.P., WARM, J.S., & ALLUISI, E.A. Effects of temporal uncertainty on watchkeeping performance. *Perception and Psychophysics*, 1966, 1, 293-299.

- SPERLING, G. The information available in brief stimulus presentations. *Psychological Monographs*, 1960, 74, (Whole No. 498).
- SPERLING, G. A model for visual memory. *Human Factors*, 1963, 5, 19-31.
- STONE, M. Models for reaction time. *Psychometrika*, 1960, 25, 251-260.
- SWETS, J.A. The relative operating characteristics in psychology. *Science*, 1973, 182, 990-1000.
- SWETS, J.A., & BIRDSALL, T.G. Deferred decision in human signal detection: A preliminary experiment. *Perception and Psychophysics*, 1967, 2, 15-28.
- SWETS, J.A., & GREEN, D.M. Sequential observations by human observers of signals in noise. In J.A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley and Sons, 1964, pp. 221-242.
- SWETS, J.A., TANNER, W.P., & BIRDSALL, T.G. Decision processes in perception. *Psychological Review*, 1961, 68, 301-340.
- TANNER, W.P., & SWETS, J.A. A decision-making theory of visual detection. *Psychological Review*, 1954, 61, 401-409.
- TAUB, H.A., & OSBORNE, F.H. Effects of signal and stimulus rates on vigilance performance. *Journal of Applied Psychology*, 1968, 52, 133-138.
- TEICHNER, W.H. The detection of a simple visual signal as a function of time of watch. *Human Factors*, 1974, 16, 339-353.
- THOMAS, E.A.C., & LEGGE, D. Probability matching as a basis for detection and recognition decisions. *Psychological Review*, 1970, 77, 65-72.
- THURMOND, J.B., BINFORD, J.R., & LOEB, M. Effects of signal to noise variability over repeated sessions in an auditory vigilance task. *Perception and Psychophysics*, 1970, 7, 100-102.
- TOATES, F.M. *Control theory in biology and experimental psychology*. London: Hutchinson, 1975.
- TREADWELL, E. The effects of depressant drugs on vigilance and psychomotor performance. In H.J. Eysenck (Ed.), *Experiments in personality: Psychogenetics and psychopharmacology*. New York: Praeger, 1960, pp. 159-196.

- TREISMAN, M., & WATTS, T.R. Relation between signal detectability theory and the traditional procedures for measuring sensory thresholds: Estimating d' from results given by the method of constant stimuli. *Psychological Bulletin*, 1966, 66, 438-454.
- UNDERWOOD, B.J. Attributes of memory. *Psychological Review*, 1969, 76, 559-573.
- VICKERS, D. Evidence for an accumulator model of psychophysical discrimination. In A.T. Welford and L. Houssiadas (Eds.), *Current problems in perception*. *Ergonomics*, 1970, 13, 37-58.
- VICKERS, D. Some general features of perceptual discrimination. In E.L. Asmussen (Ed.), *Psychological aspects of driver behavior*. Institute for Road Safety Research, S.W.O.V., Voorburg, The Netherlands, 1972.
- VICKERS, D. An adaptive module for simple judgment. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978, pp. 599-618.
- VICKERS, D. *Decision processes in visual perception*. London: Academic Press, 1979.
- VICKERS, D. Theories of confidence. In preparation, 1982.
- VICKERS, D., & LEARY, J. Criterion control in signal detection. *Human Factors*, 1982, in press.
- VICKERS, D., LEARY, J., & BARNES, P. Adaptation to decreasing signal probability. In R.R. Mackie (Ed.), *Vigilance: Theory, operational performance, and physiological correlates*. New York: Plenum Press, 1977, pp. 679-703.
- VICKERS, D., NETTELBECK, T., & WILLSON, R.J. Perceptual indices of performance: The measurement of 'inspection time' and 'noise' in the visual system. *Perception*, 1972, 1, 263-295.
- VITZ, P.C., & HAZAN, D.N. Memory during probability learning. *Journal of Experimental Psychology*, 1969, 80, 52-58.
- WARM, J.S., & ALLUISI, E.A. Influence of temporal uncertainty and sensory modality of signals on watchkeeping performance. *Journal of Experimental Psychology*, 1971, 87, 303-308.
- WELFORD, A.T. *Fundamentals of skill*. London: Methuen, 1968.
- WILKINSON, R.T. *The effects of sleep loss on performance*. Medical Research Council Applied Psychology Research Unit Report, Cambridge, No. 323, April, 1958.

- WILKINSON, R.T. Effects of up to 60 hours' sleep deprivation on different types of work. *Ergonomics*, 1964, 7, 175-186.
- WILKINSON, R.T., EDWARDS, R.S., & HAINES, E. Performance following a night of reduced sleep. *Psychonomic Science*, 1966, 5, 471-472.
- WILLIGES, R.C. Within-session criterion changes compared to an ideal observer criterion in a visual monitoring task. *Journal of Experimental Psychology*, 1969, 81, 61-66.
- WILLIGES, R.C. The role of payoffs and signal ratios in criterion changes during a monitoring task. *Human Factors*, 1971, 13, 261-267.
- WILLIGES, R.C. Manipulating the response criterion in visual monitoring. *Human Factors*, 1973, 15, 179-185.