PERFORMANCE CHANGES IN PERCEPTUAL DISCRIMINATION AND IDENTIFICATION

Jeanette Packer
B.A.(Hons.) (Adelaide)

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

Department of Psychology, The University of Adelaide
South Australia

March, 1984

Awarded 20-9-84

TABLE OF CONTENTS

Page

Page

Page

Page

CHAPTER 9 (continued)

LIST OF TABLES

TABLES (continued)

TABLES (continued)

TABLES (continued) Page

TABLES (continued)

LIST OF FIGURES

FIGURES (continued)

FIGURES (continued) Page

FIGURES Page

FIGURES (continued) Page

FIGURES (continued) Page

SUMMARY

Early studies of discrimination and identification were generally concerned with effects of discriminability on accuracy and response time. In more recent studies, other independent variables, namely bias and caution have been examined, and confidence ratings have also been recorded. While early studies concentrated on the measurement of sensory thresholds, more recent studies have emphasised the importance of the notion of a variable criterion underlying observed changes in performance.

In some previous work it has been assumed that accuracy, time and confidence each had a simple relationship to criterion values. In these studies, the aim appeared to be to arrive at a single self-regulating decision mechanism for discrimination and identification which could account for variations in accuracy, time and confidence associated with manipulations of discriminability, bias or caution.

The present thesis is aimed at examining the adequacy of a number of proposed mechanisms in the light of changes in accuracy, time and confidence which followed the manipulation of discriminability, bias and relative and overall caution. Firstly, an attempt was made to manipulate bias by variations in *a priori* probability according to either a between-subjects design or a within-subjects design in three-category tasks and in two-category tasks. Secondly, an attempt was made to manipulate bias by changes in discriminability within subjects in a two-category task. Finally, relative and overall caution were each varied by instructions according to a within-subjects design in a two-category task.

The empirical findings were compared with the predictions of an ideal observer hypothesis, in which criterion regulation is based on subjective estimates of stimulus probability, a response stabilisation hypothesis, in which the observer is assumed to adjust criteria with reference to the perceived discrepancy between cumulative and recent assessments of stimulus probability (Vickers, Leary & Barnes, 1977), an adaptation-level hypothesis, in which the observer adjusts criteria according to an average of all observations of the stimuli (Vickers & Leary, 1983), and a target confidence mechanism, in which the observer is assumed to adjust criteria in an attempt to maintain a certain target level of confidence (Vickers, 1979). The qualitative predictions were derived when each of these mechanisms was applied to each of Signal Detection Theory (Green & Swets, 1966), a random walk model of the type proposed by Stone (1960) and developed by Laming (1968) and Link and Heath (1975), and the accumulator model (Vickers, 1979).

When an attempt was made to manipulate bias by changes in a priori probability, it appeared that the observer's estimate of stimulus probabilities was a major determinant of any variation in criterion values, and that therefore, an ideal observer type of mechanism or a response stabilisation hypothesis might be the most appropriate form of criterion regulation mechanism in this situation. However, it appeared that an adaptation-level type of process might also be influencing performance independent of the other mechanisms.

When an attempt was made to manipulate bias by changes in discriminability, the major determinant of criterion regulation appeared to be an adaptation-level process. However, again there was evidence of the independent influence of another mechanism, in this case, based on the observer's estimate of stimulus probabilities.

When relative and overall caution were manipulated by instructions, the observer appeared to regulate criteria in order to maintain a given value of target confidence indicated by instructions. There was no evidence of the influence of other mechanisms in these experiments.

When these findings were considered together, it was apparent that a single criterion regulation mechanism could not account for all the changes in accuracy, time and confidence which occurred as a result of the manipulation of the three different independent variables. Instead, it appeared that any model of these changes in performance should allow for the possible independent influence of at least three adaptive mechanisms. In particular, it appeared that the observer's estimate of stimulus probability, the value of the adaptation level, and the effects of instructions or subjective costs and payoffs may separately influence the decision process, with the significance of each factor in the decision process determined by the relative importance of each independent variable, and the nature of their variation. It was suggested that these three mechanisms could be incorporated into existing decision models. For example, in terms of either a random walk or an accumulator model, it is possible that the effects of the observer's estimate of stimulus probability may influence the position of the starting point, the value of the adaptation level would determine the indifference point and hence, the rate of drift, while the effects of instructions or costs and payoffs would be reflected in the boundary values.

These findings would appear to have implications for the generally accepted notion of response bias and for the interpretation of performance changes in discrimination and identification.

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_

Jeanette Packer

March, 1984

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor, Dr. Douglas Vickers, for his invaluable advice and encouragement, and also to Dr. Ted Nettelbeck and Dr. David Caudrey, who provided generous support and assistance during Dr. Vickers' absence on study leave.

I also wish to thank Peter Barnes and Mark Brown for their expert help in computer programming, Margaret Blaber for her unfailing patience and support while typing the manuscript, Judy Fallon for her tireless efforts in the preparation of the tables and figures and Helena van Ruth for her photographic work.

CHAPTER 1

A. INTRODUCTION

The ability to discriminate differences or identify equalities among features of the environment is fundamental to most, if not all, human perceptual and cognitive activity. The existence of these abilities is assumed in all areas of psychological research and practice. For example, in psychotherapeutic techniques, the client may be asked to differentiate between his behaviour and that of a model. Similarly, in the very different field of memory studies, a subject may be asked to decide if a given item is the same as one presented before. Whether in the laboratory or in real life situations, it is difficult, if not impossible, to distinguish an aspect of human behaviour which does not involve some form of discrimination or identification.

Discrimination and identification are possible with respect to a large number of dimensions of many different kinds, ranging from unidimensional, objective scales, such as weight or length, to multidimensional, subjective scales of preference. Despite this diversity, in general it appears to be assumed that the mechanisms underlying these judgments have the same basic properties, irrespective of the sensory modality or stimulus dimension involved. For this reason, experimental studies of discrimination and identification have employed a wide range of tasks and a variety of stimuli. However, because of the importance of vision in human activity, and because visual stimuli are more conveniently manipulated and presented, the majority of these studies have employed visual stimuli, and most, though not all experiments considered in this thesis are drawn from the field of visual perception.

The assumption of a set of mechanisms common to sensory modalities and stimulus dimensions is paralleled by the use of a restricted number of experimental designs in their investigation. Of these, the three classical procedures proposed by Fechner (1860, 1966) for the empirical determination of absolute and differential thresholds are the most familiar and have been outlined in a number of standard introductions to experimental psychology (e.g. Osgood, 1953; Guilford, 1954; Woodworth & Schlosberg, 1954) as well as in some more recent texts on psychophysics (e.g. Corso, 1967; Gescheider, 1976). It is useful to begin with a brief review of some of the main classical concepts and procedures, since this thesis makes frequent reference to them. Following this, an attempt is made to summarise the general features of human performance in discrimination and identification tasks.

B. THE MEASUREMENT OF DISCRIMINATION AND IDENTIFICATION IN CLASSICAL PSYCHOPHYSICS

In all three classical procedures, the observer is presented with two stimuli which vary along a common dimension. The standard stimulus, S, is of constant magnitude, and the variable stimulus, V, may take a range of values. In the simplest procedure, the method of adjustment, there are two types of trials. On ascending trials, V starts at a value less than S, and the observer adjusts the physical magnitude of V until it is judged to be equal to S. On descending trials, V starts at a value greater than S, and is reduced in magnitude by the observer until it appears equal to S. A second procedure, the method of limits, also comprises ascending and descending trials, but allows for greater experimental control of the stimulus values experienced by the observer. On ascending trials, the value of V is increased in regular steps by the experimenter, until the observer's judgment changes from one of 'less than' to one of 'equal to' or 'greater than'. The complementary convention is adopted on descending trials. Thirdly, in the method of 'constant stimuli', several different magnitudes of V, greater than, equal to, or less than S, are paired with S in some unpredictable order, and, at each presentation, the observer is required to judge whether the variable appears greater than, equal to, or less than the standard with respect to the dimension in question.

Since results obtained by the first two methods were found to be influenced by so called errors of anticipation and habituation (Woodworth & Schlosberg, 1954; Hake & Rodwan, 1966; Corso, 1967), the majority of studies of discrimination and identification have employed one form of the method of constant stimuli (Guilford, 1954). In this method, V and S may be presented simultaneously (e.g. Henmon, 1911; Johnson, 1939; Festinger, 1943a,b) or successively (e.g. Garrett, 1922; Pierrel & Murray, 1963). This last method is the only one which does not rely on the use of three categories of judgment.

C. GENERAL FEATURES OF DISCRIMINATION AND IDENTIFICATION

In general, studies of discrimination and identification have been concerned with the effects of a small number of independent variables on two or three aspects of performance. The most common independent variable is that of discriminability, usually measured by the physical stimulus difference (V-S) between the value of the variable stimulus, V, and an actual or inferred standard stimulus, S. In addition, two other more subjective independent variables have been studied. The first is response bias, or the preference of the observer for making one rather than other possible responses. More recently, attention has been devoted to a second variable, caution, or the overall care with which the observer makes all his judgments.

The great majority of studies has been concerned with the effects of these three independent variables on three dependent measures. The dependent variable of interest in most early studies is some measure of response accuracy (i.e. the relative frequency with which the possible responses are made correctly or incorrectly). A second common measure has been the time elapsing between the presentation of the stimulus and some overt indication of the judgment made. A third, though less common dependent variable, is some measure of confidence (or the degree of certainty expressed by an observer in the accuracy of his judgment).

In the following examination of empirical features of human performance in discrimination and identification tasks, we will look first at the effects of variations in the first independent variable (that of discriminability), on each of the three dependent measures of accuracy, time and confidence. We will then look, in turn, at the effects of changes in the two other independent variables of response bias and caution.

1. Effects of discriminability

(a) Accuracy

(i) two-category judgment. A universal finding in discrimination experiments in which stimulus difference has been varied is that the proportion of correct responses is a direct function of stimulus difference (e.g. Henmon, 1906; Lemmon, 1927; Kellogg, 1931; Johnson, 1939; Festinger, 1943a,b; Pickett, 1964, 1967, 1968; Morgan & Alluisi, 1967; Vickers, Caudrey & Willson, 1971). In general, when V is much smaller than S, the probability that the observer will make a response of the form 'V>S' is low, but increases to about 0.5 as V and S become nearly equal, and continues to increase, approaching 1.0 as V becomes clearly greater than S. The curve obtained when the probability of making a response of the form 'V>S' is plotted against stimulus difference (V-S), is termed a psychometric function, after Urban (1910), and was first observed by Pierce and Jastrow (1885) to be a smooth S-shaped curve (see figure 1(a)).

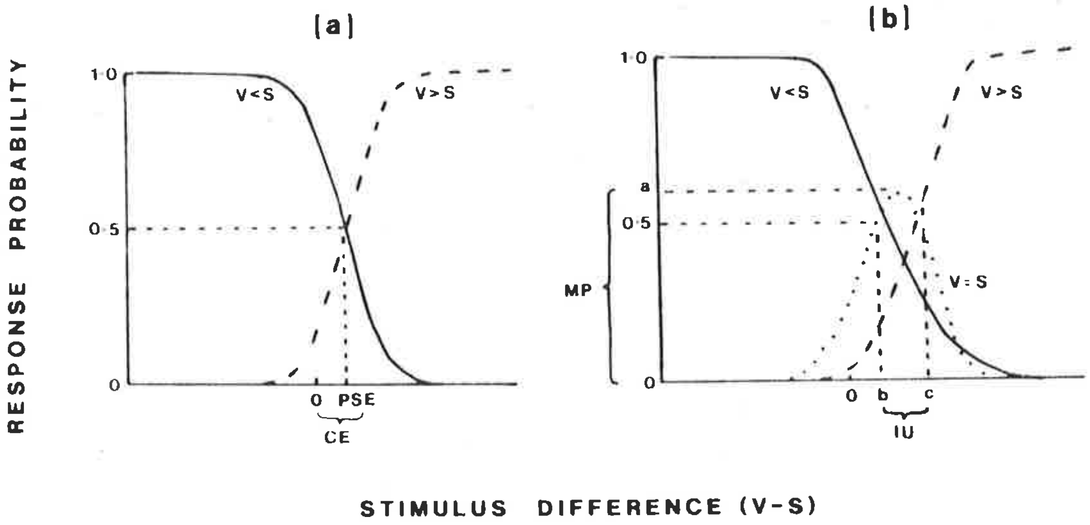


FIGURE 1.(a) Idealised psychometric functions for two-category tasks (adapted from Vickers, 1979, fig. 5). Probabilities of responses 'V>S' (broken line) and 'V<S' (solid line) are plotted as a function of stimulus difference (V-S). The difference between zero and the point of subjective equality (PSE) is a measure of constant error (CE).

(b) Idealised psychometric functions for three-category tasks (adapted from Vickers, 1979, fig. 6). Probabilities of responses 'V>S' (broken line), 'V<S' (solid line) and 'V=S' (dotted line) are plotted as a function of stimulus difference (V-S). The value, a, corresponds to the maximum probability (MP) of responses 'V=S'. The values b and c represent the values of stimulus difference at which 50% of 'V<S' and 'V>S' responses are made, respectively, while the distance between b and c represents the interval of uncertainty (IU).

Several different functions have been suggested to describe empirical psychometric curves. Originally, Urban (1910) and Boring (1917) proposed that the function should take the form of a cumulative normal ogive, and many studies have shown that this function provides a good fit to empirical data (e.g. Pierce & Jastrow, 1885; Cattell, 1893; Urban, 1910; Brown, 1910; Boring, 1917; Cartwright & Festinger, 1943; Festinger, 1943b; Guilford, 1954). Later, Thurstone (1928) suggested that the assumption that the function was a normal ogive was more appropriate when response frequencies were plotted against the logarithm of the values of the objective stimulus difference. However, Thurstone (1928) noted that this transformation of (V-S) would make little difference to the fit of the normal ogive when only a narrow range of stimulus differences was used, as is the case in most discrimination experiments. More recently, it has been argued that the cumulative logistic function gives a more precise description of the relationship between response frequencies and objective stimulus difference (e.g. Berkson, 1944, 1953; Guilford, 1954; Luce, 1959; Bush, 1963; Ogilvie & Creelman, 1968; Bock & Jones, 1968; Link, 1978a). However, all of these functions share the same sigmoidal shape, and Vickers (1979) notes that it is doubtful whether empirical data could be obtained which would differentiate between them.

(ii) three-category judgment. Only two experiments are known in which the effects of variations in discriminability have been examined using three-category tasks. In a brightness discrimination experiment, Kellogg (1931) found that the probability of making an extreme response ('greater than' or 'less than') decreased as the stimulus difference decreased. There was a concomitant increase in the probability of an intermediate response ('equal to') as the stimulus difference approached zero. When the probability of each of the three responses was plotted at each value of stimulus difference, the curves for the extreme responses were ogival psychometric functions as found in two-category tasks, while the curve for intermediate responses was bell-shaped (see figure 1(b)). More recently, Vickers (1975) confirmed these findings in a study of length discrimination.

(b) Response time

(i) two-category judgment. A second general result is the finding that, as stimulus difference decreases, the time taken to make an accurate discrimination between stimuli increases (e.g. Henmon, 1906; Lemmon, 1927; Kellogg, 1931; Johnson, 1939; Festinger, 1943a; Crossman, 1955; Botwinick, Brinley & Robbin, 1958; Pickett, 1964, 1967, 1968; Morgan & Alluisi, 1967; Pike, 1968, 1971; Vickers, 1970; Vickers et al., 1971).

In his interpretation of data from a study of length discrimination, Henmon (1906) suggested that this relationship could best be described as an arithmetic decrease in stimulus difference resulting in a geometric increase in response time. Later, Crossman (1955), who found the same relationship between discriminability and response time in a card sorting task, proposed a 'confusion function', according to which correct response times should be inversely proportional to the difference between the logarithms of the magnitudes of the two stimuli. According to this view, response times should vary with the ratio between stimulus magnitudes rather than with the difference between those magnitudes. However, Vickers (1967) showed that the predicted inverse linear relationship breaks down in those situations where the discrimination becomes very difficult, as in the studies of Birren and Botwinick (1955) and Botwinick et al. (1958). Specifically, as stimulus difference becomes very small, the rate of increase in times for correct responses decreases (Kellogg, 1931; Pike, 1968; Welford, 1968; Vickers, 1970; Vickers et al., 1971).

(ii) three-category judgment. Consistent with the above results, both Kellogg (1931) and Vickers (1975) found an inverse relationship between discriminability and times for correct extreme responses in three-category tasks. In a detailed examination of his data, Kellogg (1931) noted that, in all cases, times for accurate equality responses were equal to, or longer than, times for correct extreme responses to the smallest stimulus difference. In summarising his data, Kellogg (1931) suggested that the time curves for each category of response tended to be the inversion of the corresponding relative frequency curve. However, Vickers (1975) could not find any evidence to support this proposition in the case of equality responses.

(c) Confidence

(i) two-category judgment. A third general result is the direct relationship between subjective confidence and objective stimulus difference or discriminability. Garrett (1922) was the first to present empirical evidence concerning this relationship in a study of weight discrimination. As would be expected intuitively, when the difference between the two weights was large, observers were more confident than when the difference was small. Subsequently, Pierrel and Murray (1963) confirmed this finding for weight discrimination, while Johnson (1939) and Festinger (1943a) established that the relationship also held for length discrimination. Furthermore, Lund (1926) showed that this relationship also applied when observers were performing a closely related recognition task.

Examination of the relationship by Johnson (1939) and Festinger (1943a) suggested that plots of mean confidence in correct responses against discriminability resembled cumulative normal curves. Festinger (1943a) argued that the slight deviations from ogival form, characteristic of individual observers, were a product of differences between individuals in the frequency with which they used the various categories of the confidence scale.

(ii) three-category judgment. No data are known on the relative values of confidence for the three categories of response when discriminability is varied. Kellogg (1931) took confidence ratings only at the end of blocks of trials which contained all three categories of stimulus at each of the several levels of discriminability, and Vickers (1975) did not measure confidence.

2. Effects of manipulating response bias

(a) Accuracy

(i) two-category judgment. Within classical psychophysics, a main concern has been the determination of the value of stimulus difference at which an observer will make equal proportions of the two possible responses in a two-category discrimination task. This value of stimulus difference, the 50% threshold or the point of subjective equality (PSE), has been found to vary from an objective stimulus difference of zero if the observer is biassed towards one response or the other (e.g. Jastrow, 1888; Brown, 1914; Thurstone, 1948; Woodworth & Schlosberg, 1954). In the classical model, provision is made for the existence of bias in the form of a measure of 'constant error' (CE), defined as the difference between the observer's PSE and an objective stimulus difference of zero (see figure 1(a)).

It appears unlikely that individual observers will always come to an experiment with a complete absence of bias or preference for one response. While such a preference may vary from one observer to another or from one occasion to the next, it appears that the level of bias can be experimentally manipulated by the use of instructions or by varying the stimulus probabilities.

For example, response bias has been manipulated by the use of instructions encouraging the observer to give one response rather than the alternative. Festinger (1943b) found that, when observers, performing a length discrimination task, were instructed to avoid making any errors on 'shorter' responses, the PSE moved in a negative direction. Conversely, when observers were instructed to be as accurate as possible when making 'longer' responses, the value of the PSE moved in a positive direction.

Alternatively, bias may be manipulated by varying the *a priori* stimulus probability. However, this procedure has received much less attention in the study of discrimination. In a closely related study, Laming (1968, 1969) examined the biassing effects of changes in stimulus probability in a task in which observers were required to classify a series of vertical lines, presented individually, as 'long' or 'short'. In Laming's (1968) first experiment, the *a priori* probability of a response either increased or decreased between successive series of trials within a single experimental session and observers were informed of the variations as they occurred. The *a priori* probability took one of five values ranging from 0.25 to 0.75 in steps of 0.125. In the second experiment (Laming, 1968, 1969), the *a priori* probability was again varied between successive series of trials, but in this case, it either increased or decreased steadily from one series to the next throughout the session. The five probabilities used in this experiment were identical to those used in the first experiment. In this second study, observers were not given any information about changes in *a priori* probability. Laming (1968, 1969) found the same pattern of results in both experiments, whether or not the observers were informed of the changes in probability. As the *a priori* probability of one stimulus increased, the overall probability of making the corresponding response increased, while the probability of making the alternative response decreased. As a result, the probability of giving the response appropriate to the more probable stimulus incorrectly to the less probable stimulus, increased, while the probability of giving the alternative response to the more probable stimulus, in error, decreased. Therefore, it appears that the nature of the errors changes as the a priori probability is varied, even though the overall number of errors may remain the same.

(ii) three-category judgment. Although there have been comparatively few studies of bias in two-category tasks, there has been greater interest in the influence of bias with three categories of judgment. As with two-category judgments, bias in three-category tasks can be manipulated either by the use of instructions or by changing the stimulus probabilities. However, as before, it appears that the observer may respond according to a characteristic level of bias, even without the influence of experimental manipulations. For example, it has long been noted that, when an observer is asked to respond either 'greater than', 'equal to' or 'less than', he will search first for a difference between the stimuli and will only make an equality response if he fails to find any difference (e.g. Hayden, 1906; Angell, 1907). Therefore, the probability of making an extreme response will be high, while the probability of making an equality response will be low. This strategy seems to imply that observers have a natural tendency to avoid equality responses. However, using a weight discrimination task, Fernberger (1914a,b, 1931) showed that this attitude could be countered by the use of instructions emphasising the validity of making equality responses. Conversely, Brown (1910) found that the tendency to avoid equality responses could be strengthened by instructions. In his study of weight discrimination, observers were induced to make very few equality responses by means of instructions which stressed that it was possible to detect a difference between every pair of stimuli.

Only one study is known in which variations in a priori probability were used to manipulate the observer's bias in a three-category task. Vickers (1975) performed two experiments using a length discrimination task in which the variable stimulus could be longer than, equal to, or shorter than the standard. In addition, the variable and the standard could differ by one of five lengths so that there were eleven possible stimulus combinations. The a priori probability of equal stimuli was varied from low (0.09) in the first experiment to high (0.50) in the second experiment while, in both experiments, the five levels of discriminability for longer stimuli and for shorter stimuli were equiprobable. In this way, an attempt was made to bias the observers first away from and then towards equality responses. Vickers (1975) observed that when the a priori probability of an equal stimulus changed from low to high, there was a corresponding increase in the proportion of equality responses, as shown by an increase in the maximum probability of these responses (MP) from 0.30 to 0.73. At the same time, there was a decrease in the probabilities of both extreme responses.

(b) Response time

(i) two-category judgment. The influence of response bias on the time taken to make a response has been studied extensively using choice reaction time tasks (e.g. Hyman, 1953; Crossman, 1953; Bertelson, 1961; Lamb & Kaufman, 1965; Remington, 1971; Welford, 1971, 1973, 1975) and signal detection tasks (e.g. Laming, 1962; Jerison & Pickett, 1964; Carterette, Friedman & Cosmides, 1965; Clark, 1966; Clark, Brown & Rutschmann, 1967; Gescheider, Wright & Evans, 1968; Gescheider, Wright, Weber, Kirchner & Milligan, 1969; Williges, 1969, 1971; Pike, McFarland & Dalgleish, 1974; Pike & Dalgleish, 1976). The effects of variations in bias on response time have received much less attention in the study of discrimination. However, once more Laming's (1968, 1969) experiments in a closely related area provide some evidence on the relationship. In both of the experiments described above, Laming (1968, 1969) found that, as the a priori probability of one stimulus increased, times for the corresponding correct response decreased while times for responses to the less probable stimulus (again made correctly) increased. When this finding is considered together with the accuracy data from the experiments, it can be seen that the time taken to make a given response decreases as the probability of making that response, either correctly or incorrectly, increases. Audley (1973) considered this to be a general finding in discrimination performance and termed it the 'stimulus probability effect'.

(ii) three-category judgment. As seen above, Hayden (1906) and Angell (1907) suggested that the strategy adopted by observers performing three-category tasks commonly resulted in few equality responses being given. In addition, they noted that, since this strategy meant that equality responses would be made only after the observer had failed to find a difference between the stimuli, the time taken to make an equality response should be longer than the time taken to make an extreme response. Empirical evidence in support of this contention has been obtained in several studies of discrimination. For example, in a series of experiments on the discrimination of sound intensities and length, George (1917) found that the average times for 'greater than' responses and for 'less than' responses were 0.92 sec and 0.88 sec, respectively, while 'equal' responses took 1.0 sec on average. More recently, Kellogg (1931) found that times to make equality responses were 40-45% longer than times for correct extreme responses at the largest stimulus difference.

Similar results have been obtained in the closely related field of same-different judgments. Both Martin and Muller (1899) and Hayden (1906) observed that 'like' responses took longer than 'unlike' responses. In introspective reports, Hayden's (1906) observers indicated that " 'like' judgments have an element of uncertainty in them depending upon the absence of the criteria available in the other classes of judgments, with the consequence that they are not only longer, but also more variable" (p. 509). Cartwright (1941) also noted the longer times for equality responses in his study in the very different field of recognition memory.

However, Carlson, Driver and Preston (1934) proposed that the observed difference between the times for equality and extreme responses was a product of the bias induced by the emphasis of instructions, even though this may not have been the aim of the experiment. They suggested that the instructions used in the experiments of both George (1917) and Kellogg (1931) had implied that observers should avoid equality responses if at all possible. In order to test this explanation, Carlson et al. (1934) performed a weight discrimination experiment in which observers were trained to accept equality judgments as equivalent to extreme judgments. After this training they found no difference between times for the three responses. The same pattern of results was obtained by Fernberger and Irwin (1932) and Fernberger, Glass, Hoffman and Willig (1934) when instructions stressing the validity of equality responses were used in the related study of absolute judgments of weight.

Kellogg's (1931) results may also have been influenced by the inclusion in his experiment of a preliminary series of two-category judgments in which the observers were instructed not to give equality responses. This may have biassed observers against making equality responses in the subsequent three-category experiment. In addition, Carlson et al. (1934) suggested that so few response time measures were taken in George's (1917) study that the average measures which were discussed were unlikely to be statistically valid.

Alternatively, Kellogg (1931) suggested that the longer times for equality responses represented an extension of the commonly observed effects of discriminability. He noted that, since responses to small stimulus differences usually took longer than responses to large stimulus differences, it was possible that equality responses were slow because they were most often given when the stimulus difference was small. However, data from his experiment on brightness discrimination did not support this proposal. Kellogg (1931) found that equality responses were slow even when they were given incorrectly at large values of stimulus difference.

Vickers (1975) has examined the biassing influence of stimulus probability. He noted that, when the *a priori* probability of an equal stimulus was low, correct equality responses were made more slowly than any extreme responses, either correct or incorrect. In contrast, when the *a priori* probability of an equal stimulus was high, correct equality responses were faster than both correct extreme responses (including those made at the smallest stimulus difference) and incorrect extreme responses.

(c) Confidence

(i) *two-category judgment*. Although many signal detection experiments involve the construction of ROC curves using confidence ratings and might, in principle, allow for an investigation of the relationship between response bias and confidence, this has not been their primary aim (e.g. Pollack & Decker, 1958; Swets, Tanner & Birdsall, 1961; Green & Swets, 1966; Katz, 1970; McNicol, 1972). In the case of discrimination, no published data are known which are concerned specifically with the relationship between response bias and confidence.

(ii) three-category judgment. Even when there is no attempt to bias observers against equality judgments, it is commonly observed that equality responses appear to involve doubt more often than extreme responses (e.g. Fernberger, 1930; Kellogg, 1931; Fernberger & Irwin, 1932). In one of the earliest studies of three-category judgment, Fernberger (1919) noted that equality responses "were usually given with a lower degree of subjective assurance" (p. 155) than extreme responses. It has been suggested that the lack of confidence which observers show in equality responses may be why very few are made.

However, as Carlson *et al.* (1934) pointed out, the instructions in these studies implied that observers should avoid equality responses. Therefore, it seems possible that the response bias induced by the instructions is determining the observer's feeling of confidence. Since Vickers (1975) did not record any confidence measures in his study of the effects of *a priori* probability, the effects of bias on confidence remain unclear. Similarly, none of the other studies of three-category judgment has employed detailed measures of confidence, relying instead on introspective reports (e.g. Fernberger, 1919) or on confidence ratings taken only at the end of each block of trials (Kellogg, 1931).

3. Effects of variations in caution

(a) Accuracy

(i) two-category judgment. Besides the effects of discriminability and bias on response measures, it has been suggested that individual observers may adopt different levels of caution, and that the actual level of caution adopted can influence performance (e.g. Wollen, 1963; Pike, 1968; Vickers, 1979). However, only four studies are known in which there was an explicit manipulation of the level of caution adopted by observers performing a discrimination task. In a length discrimination experiment, Garrett (1922) manipulated caution by varying the length of time for which the stimulus was exposed before a response was required. In separate experimental sessions exposure time was either 0.2 sec or 2.0 sec, encouraging the adoption of low and high levels of caution, respectively. Under this manipulation, Garrett (1922) observed that accuracy for both responses was higher when the exposure was longer and allowed greater caution. Using similar two-category length discrimination tasks, Johnson (1939) and Festinger (1943a) manipulated caution by varying the instructions given to observers. Both Johnson (1939) and Festinger (1943a) found that, when instructions placed no particular emphasis on either the speed or accuracy of responding, the overall accuracy level was higher than when observers were instructed to respond as quickly as possible. There was a further improvement in accuracy when instructions stressed accurate responding. In a more detailed analysis of this finding, Festinger (1943a,b) noted that the precision of psychometric functions plotted for individual observers increased when instructions encouraged higher levels of caution. A further study was performed by Wilding (1974) whose observers were required to judge whether a briefly exposed point of light lay to the right or left of the centre of a screen. In contrast to previous studies, Wilding (1974) found no difference in accuracy when instructions were changed from stressing the speed of responding to emphasising accurate performance.

Although few studies have examined the effects of variations in caution on discrimination, this relationship has received more attention in the closely related field of choice reaction time. For example, Howell and Kreidler (1963) found that the accuracy of performance in a 10-choice task increased as the emphasis of instructions changed from one of speed, through 'usual' (implying no particular stress on speed or accuracy) to one of accuracy.

Similarly, Fitts (1966) noted an increase in accuracy as the speed stress of instructions was reduced in both an 8-choice task and a 4-choice task, while Hale (1969) observed a reduction in accuracy when 'speed' instructions were given in a 3-choice task. Pachella and Pew (1968) confirmed these findings in a choice reaction time experiment in which observers were induced to adopt different levels of caution in an attempt to maximise their earnings as determined by a 2 x 2 payoff matrix. The matrix dimensions were, correct vs incorrect, and, response time faster or slower than a deadline.

(ii) three-category judgment. No published data are known in which caution has been manipulated in a three-category task. However, results relevant to identification have been obtained by Link and Tindall (1971) whose observers judged whether two successively presented lines were equal in length or of different lengths. In separate sessions, observers were required to respond as accurately as possible within either a short deadline (260 msec, corresponding to speed instructions), or a medium deadline (460 msec, corresponding to instructions with no particular stress on speed or accuracy), or with no deadline at all (corresponding to accuracy instructions). In agreement with the majority of previous results, Link and Tindall (1971) found that both overall accuracy, and accuracy at each of four values of stimulus difference, were higher under the accuracy deadline condition than when the deadline was medium or short. Link (1971) and Ascher (1974) have shown that these results are also obtained when the deadline conditions in this task are varied from trial to trial.

Although there may be some problems in generalising from the findings of choice reaction time tasks and same-different judgments to the field of discrimination performance, these findings still appear to be relevant to any discussion of the notion that observers can, and do vary the degree of caution with which they respond. In all three areas of study, there are data showing evidence of changes in caution as the result of explicit experimental manipulations such as the use of instructions, deadlines or payoffs.

(b) Response time

(i) two-category judgment. The effects of changes in caution on response time measures have also been studied. In the discrimination experiments of Garrett (1922), Johnson (1939), Festinger (1943a) and Wilding (1974), response time measures were taken in addition to the accuracy measures discussed above. For example, in the length discrimination experiment described above, Garrett (1922) found that times for both responses were longer when the stimulus exposure time was long, encouraging greater caution. Johnson (1939), Festinger (1943a) and Wilding (1974) noted a similar increase in overall response times as instructions were changed from those emphasising the speed of responding to those in which there was no particular emphasis on speed or accuracy. A further increase in response times occurred when instructions stressing accuracy were used.

The studies of choice reaction time described above also included response time measures. In the experiments of Howell and Kreidler (1963), Fitts (1966) and Hale (1969) there was an unbiassed increase in response times as the stress of instructions changed from speed, through usual, to accuracy. Similarly, Pachella and Pew (1968) showed that overall response times were longer when a biassed payoff matrix favoured accurate rather than fast responding.

(ii) three-category judgment. In the studies of same-different judgment described above, Link and Tindall (1971), Link (1971) and Ascher (1974) found that the mean times for both 'same' and 'different' responses increased as the length of the deadline was increased (encouraging greater caution). Times were longest when no deadline was imposed and therefore, caution was high.

(c) Confidence

(i) two-category judgment. Only three studies are known in which an attempt has been made to examine the influence of the observer's level of caution on his confidence. These are the experiments of Garrett (1922), Johnson (1939) and Festinger (1943a). The most detailed analysis of the effects of variations in caution on confidence data was presented by Festinger (1943a) in his study of length discrimination. In this experiment, observers were asked to rate their confidence in each judgment on a scale ranging from 0 to 100. As discussed above, variations in instructions from those stressing speed to those emphasising accurate performance produced regular changes in both accuracy and response time. However, there was no corresponding change in confidence. In particular, Festinger (1943a) found that the standard deviations of ogives fitted to the empirical data relating confidence and stimulus difference for each observer did not show any consistent variation. Similarly, Johnson (1939) did not find any difference in confidence ratings given under 'speed' or 'accuracy' instructions, while Garrett (1922) found that confidence remained unchanged when stimulus exposure time was varied. The absence of any relationship between caution and confidence is surprising, since an increase in stimulus difference produces increases in both accuracy and confidence, and it might be expected that an increase in caution would also produce concomitant effects on these two dependent variables.

Ascher (1974) and Vickers (1979) suggested that this apparent inconsistency may have arisen from the particular experimental design used by Garrett (1922), Johnson (1939) and Festinger (1943a), in which manipulations of caution occurred between consecutive experimental sessions. Ascher (1974) and Vickers (1979) noted that, as a consequence, it was possible and even likely that observers would vary their employment of the confidence rating scale from one session to the next. As Ascher (1974) notes, "the criteria for confidence responses appear to be unstable and dependent upon the context in which they are produced" (p. 14). For example, in accordance with instructions to use an entire percentage scale, observers may have decided to rate their most confident response in any one session as 100, and their least confident response as 0. That is, the observers may have chosen ratings which expressed their degree of experienced confidence relative to other judgments in the same session rather than to judgments made in different sessions. Such range effects appear to be common in experiments in which conditions are varied between sessions performed by the same observers and hence in which differences are measured within observers (e.g. Poulton, 1973, 1975, 1979).

(ii) three-category judgment. Link and Tindall (1971) and Link (1971) did not take measures of confidence. However, Ascher (1974) found that confidence in both same and different responses increased with the length of the deadline and was highest when no deadline was imposed, encouraging maximum caution.

4. Effects of relationships between independent variables - discriminability and caution

Of the four possible interactions only that between discriminability and caution has received much attention, and only in connection with times for correct and incorrect responses.

(i) two-category judgment. It appears that the relationship between discriminability and response time depends on whether the response is correct or incorrect, and on whether the observer is aiming to achieve accurate or fast performance.

When discriminability is high, and speed is stressed, times for errors are generally equal to, or faster than those for correct responses. For example, Hornsby (1968) noted that when observers were instructed to respond 'as quickly as possible' in choosing which of two lamps was lit on a given trial, there was no significant difference between times for errors and correct responses. Laming (1968) used similar instructions in an easy two-category discrimination task in which a priori stimulus probabilities were equal. In this case, error response times were equal to, or shorter than times for correct responses when the inter-trial interval was short (1 msec or 64 msec). However, for longer inter-trial intervals (512 msec or 4096 msec) times for errors were longer than times for correct responses.

When discriminability is low, and instructions stress accuracy, it is commonly observed that errors take longer than the corresponding correct response (e.g. Cross & Lane, 1962; Pierrel & Murray, 1963; Pickett, 1967, 1968; Audley & Mercer, 1968; Pike, 1971). La Berge (1961) found that this relationship persisted when observers, discriminating the position of a light on a vertical line, were trained to an asymptotic level of accuracy. Similarly, in a task where observers were asked to indicate whether the right or left half of a screen was darker, Kellogg (1931) found that errors took longer than correct responses. However, he noted that the difference in times was less marked for those less cautious observers who responded more quickly overall.

This difference between observers is more apparent when instructions indicate that a compromise between speed and accuracy should be adopted.

Differentiation between fast and slow observers in the analysis of error and correct response times was first carried out by Henmon (1911). He found that, for two of his observers, errors were slower than correct responses, while for the third observer, distinguishable only by his overall speed of responding and lack of accuracy, there was no difference between error and correct response times. Similarly, Wollen (1963) and Pike (1968) found that, in the case of slow, cautious observers, times for error responses were longer than for the same response made correctly. However, they found no difference between the times for errors and correct responses in the case of faster, less accurate observers.

In a discussion of this problem, Wilding (1974) proposed that discriminability was the major determinant of error response times. In an experiment in which observers discriminated between lines varying in orientation, he found evidence for two different types of errors, those faster than correct responses when the stimuli were easily discriminated, and those slower than correct responses when the stimuli were difficult to discriminate. Wilding (1974) argued that, in addition to the direct effect of discriminability, the observer may interpret task difficulty as an indication of the relative importance of speed or accuracy. Thus, a situation in which stimuli are easy to discriminate would imply a speed emphasis, while a difficult discrimination task would imply a stress on accuracy. Therefore, the apparent distinction between fast and slow observers may be explained in terms of the perceived implicit instructional bias, which is, in turn, a product of the perceived or actual difficulty of the task. Similar results for confidence ratings have been obtained by Pierrel and Murray (1963).

This analysis is consistent with other results reviewed in this section. It implies that the relationship between discriminability and the times for correct and incorrect responses may be influenced by a number of factors. These include the perceived difficulty of the task, the emphasis in instructions on speed or accuracy (whether actual or implicit), and the level of caution adopted by the individual observer.

(ii) three-category judgment. No data are known which examine the relationship between discriminability and caution in three-category tasks.

5. Relationships between dependent variables

Although the relationships between the dependent variables of accuracy, time and confidence have been studied in some detail in two-category tasks, no data are known on these relationships in three-category tasks.

(a) Accuracy and response time

The relationship between accuracy and time has been considered when either discriminability or caution is varied. No studies are known which examine the influence of bias on the relationship between accuracy and time.

As noted by Pike (1968), variations in discriminability result in changes in response probability, with difficult discriminations producing probabilities close to chance (0.5), and easy discriminations producing probabilities close to 0 and 1 for error and correct responses, respectively. Since times for correct responses vary inversely with stimulus difference, which, in turn, is directly related to accuracy, it might be expected that there would be a consistent relationship between accuracy and time. Indeed, several studies have provided evidence for an inverse function both when signal detection tasks are used (e.g. Carterette et al., 1965; Sekuler, 1965; Gescheider et al., 1969; Pike & Ryder, 1973) and in discrimination tasks (e.g. Wollen, 1963; Pierrel & Murray, 1963; Pickett, 1967; Pike, 1968; Audley & Mercer, 1968). However, Pike (1968) noted that the inverse nature of this latency-probability function depends on the finding that times for errors are longer than times for correct responses, and, as seen above, this relationship is influenced by many factors including task difficulty, bias and caution.

Despite this problem, continuing interest has been shown in the relationship between accuracy and time because the predicted form of the latency-probability function varies between different models of the decision process (Audley & Pike, 1965). Therefore, it appeared that the function might form a useful basis for an evaluation of the models. Although several authors have thrown doubt on this form of evaluation (Koppell, 1976; Weatherburn, 1978; Weatherburn & Grayson, 1982), Pike and Ryder (1978) and Pike and Dalgleish (1982) claim that the establishment of the form of the latency-probability function must at least contribute to a test of the models.

The relationship between accuracy and response time has also been examined when caution is varied. It appears that, uninstructed, observers adopt a level of caution which they feel is appropriate to the implicit demands of the experimental task (Pachella, 1974). If this is so, any interpretation of accuracy and response time must necessarily involve an assumption of the observer's level of caution. The importance of this issue in studies which consider accuracy and response time measures has resulted in extensive investigation of the nature of the tradeoff between speed and accuracy. A brief review of this work will now be presented.

Many studies of the speed-accuracy tradeoff have used choice reaction tasks (e.g. Hick, 1952; Howell & Kreidler, 1963; Fitts, 1966; Pachella & Pew, 1968; Pew, 1969; Swensson, 1972a,b; Jennings, Wood & Lawrence, 1976; Wood & Jennings, 1976; Bach, Corlett & Dickinson, 1978; Grice & Spiker, 1979), or memory tasks (e.g. Swanson & Briggs, 1969; Lyons & Briggs, 1971;

Lively, 1972; Theios, 1972; Theios, Smith, Haviland, Traupman & Moy, 1973; Reed, 1973, 1976; Pachella, 1974; Dosher, 1976; Corbett, 1977; Corbett & Wickelgren, 1978). Relatively few studies of the tradeoff have employed discrimination tasks.

One of the earliest studies of discrimination which was specifically aimed at quantifying the speed-accuracy tradeoff was performed by Philip (1936). His observers were asked to tap each occurrence of one of two types of rhomboidal perforations which appeared on a strip of material which moved past them. The rate at which the strip moved increased from trial to trial for the first half of the experimental session and then decreased from trial to trial for the second half of the session. As a measure of accuracy, the number of taps made incorrectly was subtracted from the number of correct taps. When the resulting number, expressed as a proportion of the total possible number of correct taps, was plotted against the time taken for the strip to move past the observer, the curve was ogival. On the basis of these data, Philip (1936) suggested that the empirical curve relating speed and accuracy was a form of psychometric function, while Thurstone (1937) proposed that the relationship should be ogival on purely theoretical grounds. Further empirical support for this form of function was provided in a later study by Philip (1947) in which observers were required to discriminate between the frequencies of coloured dots which were displayed for one of five exposure times, ranging from .133 sec to .668 sec. In this experiment, ogival curves relating speed and accuracy were obtained both over all conditions, and at each exposure time. In both experiments, Philip (1936, 1947) noted that the exact shape of the curve varied between observers in an apparently random way.

More recently, Vickers _et al._ (1971) observed differences in the tradeoffs adopted by individual observers who were asked to indicate which was the more frequent of two flashing lamps. In this case, data from all observers were well described by a single curve, with the individual observers lying at different points on the curve. However, in contrast to previous results, Vickers _et al._ (1971) found no clear evidence for an ogival function, although the inverse nature of the relationship was obvious.

In addition to variations in the tradeoff functions obtained from different individuals, it is commonly found that the slope of the function changes with any variation in discriminability. For example, in the discrimination experiment described above, Philip (1947) found that the slopes of tradeoff functions plotted for each observer increased as the proportions of dots of different colours became more similar. Many other authors including Swensson (1972a,b), Lappin and Disch (1972a,b, 1973), Thomas and Myers (1972) and Harm and Lappin (1973) have found evidence for an increase in the slope of the function relating speed and accuracy as task difficulty increases. Lappin and Disch (1972a,b, 1973) and Thomas and Myers (1972) considered the well supported finding that tradeoff functions differ at different levels of discriminability together with previous data showing characteristic differences in the functions of individual observers, and concluded that this implied that any measure of the tradeoff should include an indicator of sensitivity, separate from the actual criterion adopted by the observer.

Further detailed study of the form of the function relating response time and accuracy has been undertaken. For example, Swensson (1972a) and Lappin and Disch (1972a,b, 1973) assessed the fit of several functions to the data of Schouten and Bekker (1967), Taylor, Lindsay and Forbes (1967), Pachella, Fisher and Karsh (1968), Pew (1969), Pachella and Fisher (1969) and Swanson and Briggs (1969). Empirically, the tested functions were as good as each other, so it was impossible to choose between them on this basis. Therefore, it appears that any choice will have to depend on other considerations, such as the specifications provided by the most adequate theoretical account of discriminative performance overall, or, at least the best theoretical explanation of the speed-accuracy tradeoff. The role of different theories in the determination of the form of the relationship will be discussed in Chapter 2.

Although it remains difficult to quantify the speed-accuracy tradeoff, it seems to be emerging as an essential concept in the study of discrimination. Indeed, Pachella (1974) and Wickelgren (1977) suggested that an examination of the exact tradeoff function which is operating in a given experiment should be carried out before response time and accuracy measures are interpreted. On the other hand, Schmitt and Scheirer (1977), Kantowitz (1978) and Dosher (1979) argued that experimental measures of response time and accuracy may be useful even when the exact tradeoff function cannot be specified.

(b) Accuracy and confidence

The relationship between accuracy and confidence has also been investigated. For example, several researchers including Williamson (1915), Jersild (1929) and Pierrel and Murray (1963) noted that, over a wide range of experimental tasks, and at different levels of discriminability, "the degree of confidence given is largely an individual matter, and is not a very reliable index of objective accuracy" (Garrett, 1922, p. 68). Within the data for individual observers, Henmon (1911) found that there was a direct relationship between confidence and accuracy in some, though not all cases. More recently, Audley and Mercer (1968) found that, when discriminability was constant, all observers showed a direct relationship between confidence and accuracy. Both sets of results are consistent with Trow's (1923) observation that, although highly confident decisions were more likely to be correct, observers who were more confident overall were not more likely to be accurate. In addition, Hertzman (1937) noted this trend in the confidence ratings of individual observers in a related study of memory, while, in contrast to previous studies, Hoge (1970) found no evidence of any relationship between confidence and accuracy in a complex decision task. In a discussion of these studies, Adams and Adams (1961) suggested that the finding of a consistent, direct relationship between confidence and accuracy in many experiments implied some validity in the confidence rating as a measure of performance.

Another related finding, which has been widely reported, is that an observer's confidence in his judgments tends to underestimate their accuracy (Pierce & Jastrow, 1885; Fullerton & Cattell, 1892; Henmon, 1911; Johnson, 1939; Festinger, 1943b; Adams, 1957). However, Howell (1967, 1970, 1971) found evidence for overconfidence in observers' predictions of their performance on a dart throwing task, while more recently, Fischhoff, Slovic and Lichtenstein (1977) and Koriat, Lichtenstein and Fischhoff (1980) found that their observers were overconfident when rating their answers to general knowledge questions. These data suggest that the finding that confidence underestimates accuracy may be restricted to simple perceptual judgments.

(c) Confidence and response time

The relationship between confidence and response time has also been studied. For example, in length discrimination tasks, Peters (1910) and Henmon (1911) found that, when the levels of discriminability were considered together, there was an inverse relationship between an observer's confidence in his judgment and the time that he took to respond. The same relationship between confidence and response time has been observed by Seward (1928) in a pattern recognition task, by Pierrel and Murray (1963), whose observers were asked to decide which of two weights was heavier, and by Emmerich, Gray, Watson and Tanis (1972), in an experiment on auditory signal detection.

Johnson (1939) suggested that the relationship between confidence and response time was logarithmic, noting that "as doubt or uncertainty increases arithmetically, judgment time increases geometrically" (p. 50), and proposed a descriptive equation

$$T = a(10)^{b(1-c)} \qquad (1)$$

where T is response time, c is confidence, a is the intercept and b is a constant. Alternatively, following a study of the inclination of pairs of lines, Volkmann (1934) suggested that the relationship was best described by a hyperbolic function of the form

$$T = \frac{d}{2c-1} + b \qquad (2)$$

where T, b and c have the same meanings as in equation (1) and d is a constant.

In contrast to these findings, it appears that in some situations, confidence may vary directly with response time. For example, Irwin, Smith and Mayfield (1956) had observers perform what they called an 'expanded judgment' task which was designed as an analogue of the hypothesised psychophysical decision process. In this task, observers were presented with a set of cards on each of which was printed a number. After the presentation of a given sample of cards, observers were required to estimate whether the mean of the pack of 500 cards from which the sample had been chosen, was greater or less than zero. They found that confidence ratings were higher after 20 cards had been viewed than after only 10 cards. This result was also demonstrated by Irwin and Smith (1957) and Little and Lintz (1965) using equivalent tasks, while a similar result was obtained in an experiment by Geller and Pitz (1970) in which confidence in probability estimates was found to increase during the course of a random sequence of events occurring with unequal probabilities.

From the review of the literature which is summarised in table 1, it is apparent that there are many areas which have received little or no attention when considering the effects of the independent variables of discriminability, bias and caution as measured by accuracy, time and confidence. Overall, few studies of three-category tasks have been undertaken and even studies of two-category tasks have tended to concentrate on the manipulation of a single independent variable as measured by at most two dependent variables. In addition, it is clear from the table that very few studies have considered more than one independent variable at a time, while relationships between response measures have been largely overlooked. As a consequence, it appears that there are many areas in which research could usefully be carried out.

DEP INDEP	a	t	c	at	ac	tc	atc
			TWO-CATEGORY TASKS				
D	27	18	5	-	15	-	-
B	7	2	-	-	-	-	-
C	4	4	3	20	-	-	-
DxB	-	-	-	-	-	-	-
DxC	-	-	-	15	-	-	-
BxC	-	-	-	-	-	-	-
DxBxC	-	-	-	-	-	-	-
			THREE-CATEGORY TASKS				
D	2	2	-	-	-	-	-
B	7	8	4	-	-	-	-
C	-	-	-	-	-	-	-
DxB	-	-	-	-	-	-	-
DxC	-	-	-	-	-	-	-
BxC	-	-	-	-	-	-	-
DxBxC	-	-	-	-	-	-	-

TABLE 1. The number of studies of discrimination and identification which have examined the effects of any of the independent variables of discriminability (D), bias (B) and caution (C) on any of the dependent measures of accuracy (a), response time (t) and confidence (c). Each entry represents the number of studies in which either two-category or three-category tasks were used to investigate the effects of the specified independent variable or variables on the specified dependent variable or variables. For example, there have been 27 studies using two-category tasks which have examined the effects of discriminability on accuracy. A dash indicates that no studies are known which fall in this category.

CHAPTER 2

A. INTRODUCTION

This chapter begins with an outline of the main theoretical models of discrimination and identification and an examination of certain features common to these models. Among these, the notion of a criterion, responsive to changes in stimulus parameters, payoffs and instructions, appears as crucial. A number of mechanisms for the regulation of criteria in these models, based on the influence of these factors, is then described.

B. MODELS OF DISCRIMINATION AND IDENTIFICATION

There are several possible classifications of the models which describe discrimination and identification. When the dependent variables of accuracy, time and confidence are being considered, a useful classification arises from the distinction between models in which the observer makes his decision on the basis of (a) a single observation, (b) a fixed sample of n (>1) observations or (c) a sample of n observations where n may vary from trial to trial. The models below are considered in this order. As this classification suggests, the models are ordered in terms of the increasing flexibility of their accounts of response time.

1. The classical model

Within classical psychophysics, a primary interest has been the determination of sensory thresholds. Fechner (1860, 1966) defined the threshold as a boundary value on the stimulus dimension which separates the stimuli which elicit one response from the stimuli which elicit a different response or no response. Although the majority of early experiments investigating threshold values studied absolute judgments, the concept of a threshold has since been generalised to apply to many types of task. In the case of discrimination, the threshold is defined as the value of stimulus difference, (V-S), which is discriminated accurately on a fixed proportion, such as 50% or 75%, of presentations.

Most of the early studies of sensory thresholds used one of the classical procedures outlined above. Commonly, data collected according to any of the three procedures have been examined by means of graphs showing the probability of a given response as a function of the value of stimulus differences (V-S) (e.g. Stevens, 1951; Guilford, 1954; Woodworth & Schlosberg, 1954). Since discrimination should be most accurate when the difference between the stimuli is large, and least accurate when the stimuli are very similar, response probability should be a positive, increasing function of (V-S). Furthermore, if perfect discrimination were possible, the curve relating response probability and stimulus difference should show an abrupt discontinuity when (V-S) takes a non-zero value and discrimination becomes possible. Below this point correct responses would never occur, while above this point all discrimination judgments would be accurate. In this case the function will be step-like, as seen in figure 2.

However, as seen above, even the earliest studies found no evidence of any discontinuity in empirical functions (e.g. Pierce & Jastrow, 1885; Brown, 1910). Jastrow (1888) was the first to suggest an explanation for the continuous nature of the psychometric function. He attributed the shape of the curve to the properties of the stimuli, proposing that the sensory representation of a given stimulus was not constant, but varied from one judgment to the next. It is commonly assumed that the combination of multiple independent and random sources of variation results in the sensory representation of V and S being best approximated by two

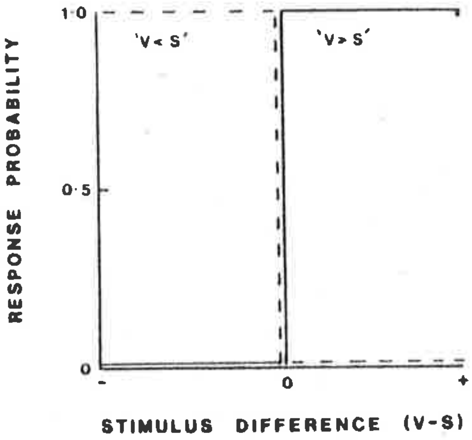


FIGURE 2. The step psychometric function which would be expected if the observer, in a two-category task, were able to discriminate perfectly. The probability of responses 'V>S' is shown by the solid line and the probability of responses 'V<S' is shown by the broken line.

normal distributions with equal variances and mean values directly related to the objective magnitudes of the stimuli (e.g. Jastrow, 1888; Cattell, 1893; Boring, 1917; Tanner & Swets, 1954). Thus, each judgment will be a function of the difference between the momentary values of V and S on that trial. Accordingly, over a series of trials, the distribution of stimulus differences (V-S) will also be normal with a mean $(\overline{V-S})$ equal to $(\bar{V}-\bar{S})$.

Based on this representation of the stimuli, Urban (1910) proposed that, when V is objectively greater than S, the mean, γ $(=(\bar{V}-\bar{S}))$, of the distribution of stimulus differences will be positive and therefore it will lie above the threshold value of zero (see figure 3). Thus, the probability, Φ, of making a response 'V>S' to a given value of (V-S) will be equal to the area under the curve to the right of the threshold (the shaded area, p, in figure 3). As shown by Boring (1917), changes in γ, the mean of the distribution of stimulus differences, will produce corresponding changes in the probability Φ,

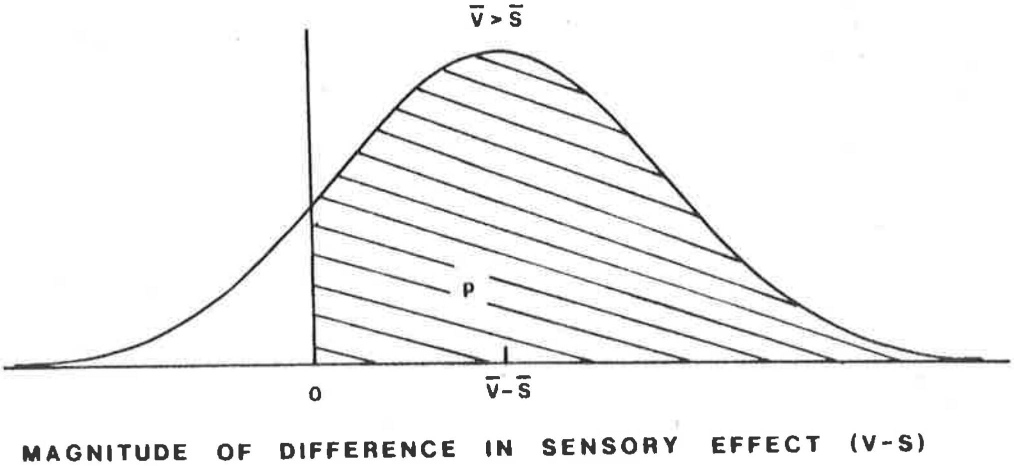


FIGURE 3. Probability density distribution of (V-S) when $\bar{V}$ is objectively greater than $\bar{S}$. The mean of the distribution, $(\bar{V}-\bar{S})$ (=γ) lies above zero. The probability, Φ, of a response 'V>S' is represented by the shaded area, p, to the right of the threshold value of zero.

of giving a response 'V>S', and the resulting function relating Φ and γ takes the form of a cumulative normal ogive. An equivalent function may be derived for the case when V is objectively less than S and therefore where γ is negative and Φ is the probability of a response 'V<S'. The proposal, commonly termed the phi-gamma hypothesis, was presented as the basis for the ogival nature of the psychometric function relating response probability and stimulus difference (e.g. Urban, 1910; Boring, 1917; Thurstone, 1928; Guilford, 1954).

The other assumptions of the model appear to pose problems when empirical findings are examined. For example, it is assumed that each response is based on a single observation which is classified in comparison to a cutoff in an error free way, and that the classification of each observation takes the same time. Therefore, the model cannot account for any variations in response time. More particularly, the model cannot explain the systematic variations which occur as a result of changes in discriminability from trial to trial. Problems for the model also arise when the speed-accuracy tradeoff is considered, since, if it is assumed that each decision is based on only one observation, there is no way of accounting for the apparent increase in discriminative capacity which is associated with taking a longer time to respond. In addition, although bias may be explained in terms of the measure of CE, there is no theoretical counterpart to explain why such variations should arise. Similarly, there is no mechanism within the model to account for any changes in performance due to a change in caution. Finally, since the model incorporates a single threshold there is no possibility of accounting for three-category judgments.

In the light of these problems, modifications to the model have taken the form of successively questioning each of the assumptions. It appears that most of the problems result from the assumption of a single observation which is classified in relation to a fixed threshold. As a consequence of this, neither changes in discriminability nor changes in the observer's attitude can be explained. The second main model, Signal Detection Theory, was designed to overcome these problems, especially with reference to detection tasks.

2. Signal Detection Theory

(a) Two-category judgment

Detailed discussions of Signal Detection Theory have been presented in several books and articles devoted to this topic (e.g. Green & Swets, 1966; Hake & Rodwan, 1966; McNicol, 1972; Egan, 1976). However, since there are frequent references to the constructs of Signal Detection Theory in this thesis, a brief account of the theory will be presented in order to explain its application to discrimination.

Signal Detection Theory was formulated in an attempt to overcome the problems which arose from the notion of a fixed threshold value in the classical model. Since absolute thresholds had been the focus of most of the experimental work in classical psychophysics, Signal Detection Theory concentrated on the detection problem, in which the observer is presented with stimuli (signals) which are weak compared with the background against which they are presented (noise). As with the classical model, Signal Detection Theory includes the assumption that the sensory effect of both signal (s) and noise (n) varies according to normal distributions, so that the detection situation may be represented by two normal distributions, one showing the sensory effect of noise alone and the other showing the sensory effect of the signal added to the sensory effect of noise. The signal + noise distribution will always lie to the right of the noise distribution along the dimension of sensory effect.

In contrast to the classical model, Signal Detection Theory introduced the notion of a variable criterion or cutoff which is adopted by the observer and which determines the decision made. According to Signal Detection Theory, the observer takes an observation of the sensory events occurring within a fixed, usually short, interval of time and makes a decision on the basis of the position of this observation relative to the criterion or the position of the cutoff on the dimension of sensory effect. All observations lying above the criterion value will be judged as occurrences of the signal, while observations lying below the criterion will be treated as noise. Consequently, responses may fall into one of four categories, a hit (the observer responds 'signal' when a signal occurred), a miss (the observer does not respond 'signal' when a signal occurred), a false alarm (FA, the observer responds 'signal' when no signal was presented) and a correct rejection (CR, the observer judges that there was noise alone when a signal was not presented).

Signal Detection Theory has been used extensively in the analysis of detection tasks (e.g. Tanner, 1956, 1958; Clarke, 1960; Egan, Greenberg & Schulman, 1961; Carterette & Cole, 1962; Eijkman & Vendrick, 1963). However, fewer studies have used Signal Detection Theory to explain data from discrimination tasks (e.g. Treisman & Watts, 1966; Bonnel & Noizet, 1979). In the case of two-category discrimination, it is assumed that the two stimulus situations (V>S and V<S) are represented by normal distributions of sensory effect lying along a continuum of sensory effect as in the classical model. For example, in a two-category length discrimination task, the distributions will take the relative positions as shown in figure 4. Rather than a fixed threshold value, Signal Detection Theory assumes that the observer will adopt a cutoff, x_c, which may be greater than, less than, or equal to zero. (When the cutoff is at zero, the model is essentially the same as the classical model.) Each observation of the stimuli will then be compared with the cutoff, and a response will be made on the basis of the position of the observation to the right or left of the cutoff. When the observation lies above the cutoff, the observer will respond 'V>S', while an observation lying below the cutoff will produce a response 'V<S'.

As in the detection situation, this decision process gives rise to four response categories which are commonly labelled in the same way as in detection tasks for the sake of convenience. The choice of which stimulus is to be denoted as the signal is arbitrary. According to the convention

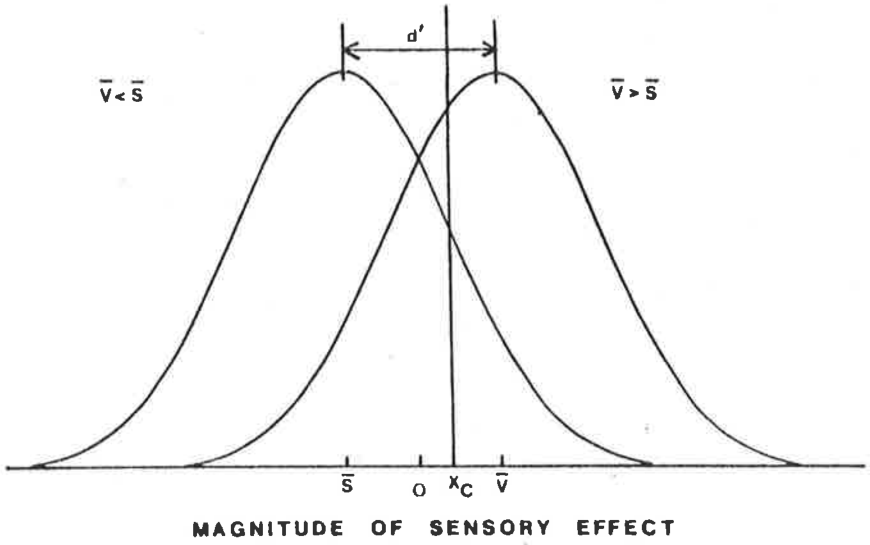


FIGURE 4. Probability density distributions of stimulus difference (V-S) appropriate to Signal Detection Theory for two-category tasks. The distribution with mean less than zero represents the case where $\overline{V}<\overline{S}$ and the distribution with mean above zero represents the case where $\overline{V}>\overline{S}$. The distance between the means of the two distributions, d', represents discriminability, while bias may be varied according to the position of the cutoff, x_c, above or below zero. The location of x_c above zero indicates a bias towards responses 'V<S' while the location of x_c below zero indicates a bias towards responses 'V>S'.

adopted in this thesis the categories are, a hit (the observer responds 'V>S' when V is objectively greater than S), a miss (the observer responds 'V<S' when V is objectively greater than S), an FA (the observer responds 'V>S' when V is objectively less than S) and a CR (the observer responds 'V<S' when V is objectively less than S).

In the analysis of detection and discrimination performance according to Signal Detection Theory, the inclusion of the notion of a variable criterion means that the observer's bias towards one response or the other can be distinguished from his sensitivity. In the terms of the model, sensitivity, or discriminability, d', is defined as the difference between the means of the two distributions of sensory effect as shown in figure 4, while bias is indicated by the location of the cutoff, x_c, above or below zero. When the cutoff is below zero the observer is biassed towards responding 'V>S'. Alternatively, when the cutoff is above zero the observer will show a preference for responding 'V<S'.

More particularly, the notion of a variable criterion provides a means whereby the effects of manipulating bias can be measured. For example, in the experiments discussed above, Laming (1968, 1969) varied bias by changing the a priori stimulus probabilities. In terms of the model based on Signal Detection Theory this manipulation of stimulus probability would be interpreted as inducing changes in the value of the criterion adopted by the observer. For example, an increase in the probability of a stimulus V>S would encourage the observer to adopt a lower criterion and hence to respond 'V>S' more often. This prediction of the model receives support from Laming's (1968, 1969) finding that observers made a higher proportion of responses which were appropriate to the more probable stimulus. Alternatively, the criterion value may change due to bias induced by the use of payoff matrices favouring one response rather than the other.

(b) Three-category judgment

Following Greenberg's (1965) modification to Thurstone's (1927a,b) law of comparative judgment, Treisman and Watts (1966) and Olson and Ogilvie (1972) suggested a model for three-category judgments which is expressed in terms of Signal Detection Theory.

In these models it is suggested that the observer adopts two cutoffs, x_g and x_l, on the distribution of stimulus difference, where x_g is greater than x_l. When the value of the observation of stimulus difference exceeds

x_g, the observer responds 'greater than', when the observation falls below x_l he responds 'less than' and when the observation falls between x_g and x_l he responds 'equal to' (see figure 5). According to this approach, the psychometric functions for the three responses should resemble those obtained by Kellogg (1931) and Vickers (1975) as discussed above. In addition, in this formulation, the first observation always gives rise to one of the three possible responses and therefore, provided that the time taken to classify the observation is independent of the size of the stimulus difference, response times for the three categories of judgment will be identical. However, empirical studies by Kellogg (1931) and Vickers (1975) have shown that the times for the three responses may differ markedly, with the times for extreme responses varying inversely with stimulus difference.

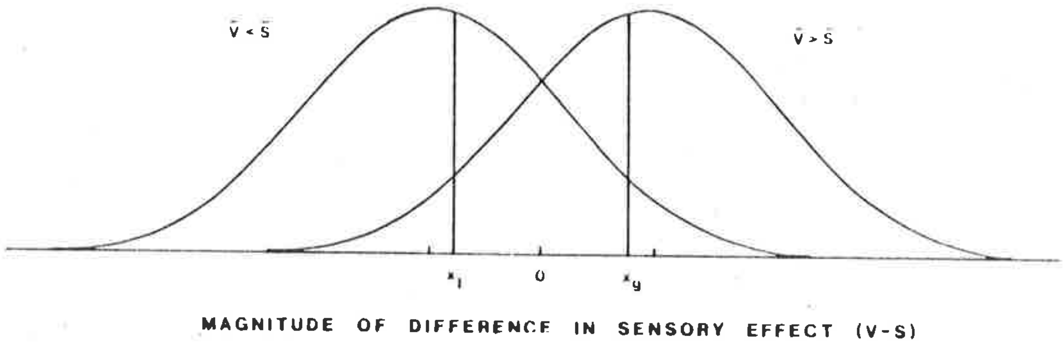


FIGURE 5. Probability density distributions of stimulus difference (V-S) appropriate to Signal Detection Theory for three-category tasks. The distribution with mean less than zero represents the case where $\overline{V}<\overline{S}$ and the distribution with mean above zero represents the case where $\overline{V}>\overline{S}$. When observations fall beyond x_l or beyond x_g the responses 'V<S' and 'V>S' will be made, respectively. When observations fall between x_l and x_g a response 'V=S' will be made.

(c) Extensions of Signal Detection Theory to account for response time and confidence

Although the model can explain the variations in response probability resulting from the manipulation of response bias in both two-category and three-category tasks, as it stands, Signal Detection Theory is not able to account for any changes in response time or confidence. However, several modifications have been suggested to enable the model to do so.

(i) response time. One possible extension of the model to account for response time is the inclusion of latency functions as proposed by Gescheider et al. (1968), and extended by Gescheider et al. (1969) and Pike (1973). The general Signal Detection model assumes that the observer responds on the basis of the information from a single observation. According to a latency function hypothesis, the time taken to classify this observation, and hence, the time taken to make a response, depends on the distance of the observation from the cutoff, with an observation which falls closer to the cutoff requiring longer to classify. The time taken to classify an observation decreases, according to the specified latency function, as the distance between the observation and the cutoff increases.

The inclusion of one or more latency functions provides a means whereby the theory can account for variations in response time resulting from the manipulation of discriminability, bias or caution. For example, the increase in response time which occurs as stimulus difference decreases may be explained if it is assumed that, when the stimulus difference is small, observations tend to lie close to the cutoff, and that the distance between the observation and the cutoff tends to increase with stimulus difference. Changes in response time due to variations in bias may be implemented by the inclusion of a different, probably asymmetrical latency function for each response, while changes in response time due to different levels of caution would be associated with symmetric changes in the latency functions for both responses.

Another modification to the model to enable it to account for response time data was presented by Swets and Birdsall (1967). They suggested that, rather than basing his decision on a single observation, the observer takes several observations, using his knowledge of signal strength, signal probability and costs and payoffs to predetermine the number of observations on which he bases his decision on any one trial. In this case, the decision is based on the average of the positions of the observations. However, this mechanism would be unlikely to produce the systematic variations in response time which are observed when discriminability is varied randomly from trial to trial.

(ii) confidence. Confidence measures may be accommodated by the model if an hypothesis, suggested by Audley (1960), is adopted. In general terms, Audley (1960) proposed that the confidence which an observer expresses in a judgment is an inverse function of the time taken for a response. This hypothesis can be combined with either of the extensions of the model which account for response time. However, since the latency function hypothesis appears to provide a more adequate account of existing empirical results, only this alternative will be considered here. According to the latency function hypothesis, response time increases as the distance between the observation and the cutoff decreases. Therefore, following Audley (1960), confidence will decrease as the distance between the observation and the cutoff decreases. In other words, confidence will vary directly with the distance between an observation and the cutoff. However, the ad hoc nature of the choice of a latency function determines any specific predictions about confidence.

Several other suggestions have been made which enable Signal Detection Theory to account for variations in response time and confidence. Perhaps the most detailed proposals are the optional-stopping models suggested by Pike (1973) and Pike *et al.* (1974) which will be discussed below.

3. Models involving multiple observations and optional-stopping rules

(a) General assumptions and features of models for two-category judgment

Several modifications to the model based on Signal Detection Theory have been suggested on the basis of the proposal that the observer may take more than one observation before making a decision. For example, Crossman (1955) suggested that the observer might make a series of y observations of the stimulus difference and respond according to an average of the observations. Crossman (1955) argued that, if it is assumed that observations are taken at a steady rate and that each observation takes the same time, then d' should increase as a function of the square root of y. Therefore, unlike Signal Detection Theory, Crossman's (1955) model can account for variations in response time as well as in accuracy. Empirical evidence in support of this relationship between d' and y has since been presented by Taylor *et al.* (1967) in a re-analysis of data from auditory and visual discrimination experiments performed by Schouten and Bekker (1967).

This model has difficulty accommodating the regular changes in response measures, especially the covariation of time and errors, which occurs when discriminability is varied. In particular, the model does not include any rationale for the adjustment of y from trial to trial in order to account for the empirical findings when discriminability is varied randomly between trials.

In view of these difficulties, recent models of discrimination have incorporated an additional assumption, that of optional-stopping, whereby the observer may respond after a single observation, if the evidence in favour of a given response is sufficiently strong, or may continue to take observations until sufficient evidence is obtained. In addition, all of these models share most or all of the following assumptions or features. Firstly, unlike Signal Detection Theory, these models do not assume that sensory effect varies only from trial to trial. Instead, following Thomson (1920), it is assumed that the sensory effect varies from one instant to the next within a single trial. Thus, on any one trial, the observer may take a series of observations from the distribution of sensory effect, each of which is likely to differ in the amount of evidence that it provides in favour of a given response. As in Crossman's (1955) model, it is assumed that observations occur at a steady rate and are of constant duration. Secondly, many of these models assume that the observer stores the information obtained on each observation, and that some aspect of the data from all observations is taken into account when a decision is made. Thirdly, the models include the notion of a criterion amount of information which is required before a response can be made. Fourthly, in contrast to Signal Detection Theory, the models incorporate the possibility that each response may have a different criterion value and that these criteria may vary independently. Fifthly, it is assumed that, when stimuli are easy to discriminate, the majority of observations of the stimulus difference will favour the correct response. Therefore, it is predicted that when discriminability is high, the criterion amount of information for a response will be collected after fewer observations, and as a result, response times will be short. At the same time, it would be expected that accuracy will be high, since it is unlikely that observations will favour the incorrect response. A sixth common feature of the model is that variations in response bias are implemented by differential movements in the criterion values. For example, in a two-category discrimination task, it is assumed that the preferred response has a relatively low criterion when compared with the non-preferred response. Therefore, it is predicted that the lower criterion for the preferred response will tend to be satisfied after fewer observations, and that, as a consequence, times for this response will be shorter. In addition, the preferred response is likely to be made accurately to the appropriate stimulus. On the other hand, the preferred response may also be given in error to the alternative stimulus because the lower criterion is more likely to be satisfied in error. The seventh assumption shared by all of the models is that changes in the observer's level of caution are implemented by a matching increase or decrease in the criteria for both responses, with high levels of caution corresponding to high criteria and low levels of caution corresponding to low criteria. Thus, when caution is high responses will be slower and more accurate than when caution is low.

(b) Specific models for two-category judgment

Although the models share a number of common features, each model can be distinguished by the nature of the information remembered from each observation, and the precise formulation of the criterion for a decision.

In a proposal resembling an earlier model by Cartwright and Festinger (1943), Swets and Green (1961) suggested that an observer may adopt two cutoffs, rather than the single cutoff assumed in Signal Detection Theory. In this case, if an observation falls beyond either cutoff the appropriate response is made. If an observation falls between the cutoffs the observer continues to take observations until he obtains one which is more extreme than one of the cutoffs, at which point the corresponding response will be made. Therefore, as in Crossman's (1955) model, d' should increase as a function of $\sqrt{y}$. However, unlike Crossman's (1955) model, this model has no memory, since each decision is based on evidence from only the last observation. In addition, the magnitude of the observation is irrelevant to the decision.

A second model, proposed by Audley (1960), is commonly called the runs model. According to this model, the observer makes a response when a run of k observations favouring a particular response occurs, uninterrupted by observations favouring the alternative response. Thus, the model has a partial memory, discarding the evidence from any runs which do not satisfy the criterion and registering only the single critical run of observations which triggers a response. As in the earlier models, on each observation the observer notes only which response is favoured, and does not take account of the magnitude of the evidence obtained from the observation.

A third model, the recruitment process of La Berge (1962), involves a more complete memory for the sampled information. According to this model, observations may be positive or negative, favouring one of the possible responses, or neutral, favouring neither response. The criterion in this process is the attainment of a predetermined total number of observations in favour of a particular response, irrespective of the number of neutral observations or observations favouring the alternative response which intervene among the sampled series. In this way, the recruitment model takes account of all of the observations which are made prior to a response. However, as with the earlier models, the magnitude of the observations is ignored.

A fourth model of this general form is the random walk model which was first presented by Edwards (1965). In this model a response is made when the number of observations in favour of one response exceeds the number favouring the alternative response by a given critical number. All observations are considered when making a decision, but their magnitude is not taken into account.

Another version of the random walk model which does take magnitude into account was proposed by Stone (1960) and Laming (1968) and has been developed by Link (1975, 1978a,b) and Link and Heath (1975). In this version, it is assumed that each observation provides evidence in favour of a particular response, and that the magnitude of this evidence may vary from observation to observation. A response is made when the total of evidence in favour of one response exceeds the total of evidence in favour of the alternative response by a critical amount.

A fifth model, which is closely related to the modified random walk model, is the accumulator model, proposed by Vickers (1970, 1978, 1979). In this process, it is assumed that, on each observation, the observer notes both which response is favoured and the magnitude of the evidence in favour of that response. A response is made when the total amount of evidence in favour of one response reaches a criterion value specific to that alternative.

So far, evaluation of models has been conducted in terms of response probabilities and times, and the evidence favouring the various alternatives has been reviewed by Audley (1970), Rabbitt (1971), Vickers et al. (1971), Pachella (1974), Wilding (1974) and Link (1975, 1978a, 1979).

The main differences between the various models are concerned with the prediction of the relative times for correct and incorrect responses, and with the changes in the shape of the response time distributions as a function of discriminability.

For example, Vickers et al. (1971) and Wilding (1974) have argued that neither a runs nor a recruitment process can account for the interaction between discriminability, caution and the relative times for correct and incorrect responses obtained in the experiments of Kellogg (1931), La Berge (1961), Pickett (1967, 1968), Laming (1968) and Pike (1971).

Again, Vickers et al. (1971) and Wilding (1974) pointed out that measures of the shape of the response time distributions for the runs and recruitment models (standard deviation, skew and kurtosis) do not correspond to empirical findings (e.g. Vickers et al., 1971; Wilding, 1974, 1978; Vickers, 1979).

At the same time, the evaluation of models by means of the comparison of empirical and theoretical distributions has been questioned. For example, Taylor (1965) and Wilding (1978) noted that empirical distributions may be influenced by factors other than the simple statistical variation in the number of observations required to satisfy a criterion. Moreover, Ratcliff (1979) has pointed out that the moments of a response time distribution are very susceptible to influence by outliers. Thus, it may sometimes be misleading to use direct quantitative comparison between empirical and theoretical measures as the basis for an evaluation of the models.

(c) Specific models for three-category judgment

A modification of the random walk model to include equality judgments, based on a model for same-different judgments proposed by Nickerson (1969, 1971), is outlined by Vickers (1975). In this 'random walk and clock' model it is assumed that the observer takes a series of observations from the normal distribution of stimulus differences and sorts them into separate stores according to whether they are positive or negative. The observer continues to take observations either until the number in one store exceeds the number in the other store by a critical amount, whereupon the appropriate response of 'greater than' or 'less than' is made, or until a set time, t, has elapsed, when a response 'equal to' is given. This proposal is based on the existence of a neural clock which has been discussed previously by McGill (1963), Sekuler (1965) and Bindra, Williams and Wise (1965).

Simulations of this model by Vickers (1975) showed that the psychometric functions for the three response categories correspond to those obtained by Kellogg (1931) and Vickers (1975). However, contrary to the findings of these experiments, the model predicts that all equality judgments will be made with response time t, independent of the stimulus difference. In addition, since equality responses are made only when an extreme response cannot be given within the time t, all equality responses should be slower than extreme responses. This prediction is not supported by data from the experiments of Carlson et al. (1934), Cartwright (1941) and Vickers (1975) discussed above, in which observers were biassed towards extreme responses.

One possible solution to these problems was proposed by Vickers (1975), who suggested that the clock may not be set at a constant time t, but may vary at random. As a result, it would be possible for the time taken to make an equality response to be shorter than the time for an extreme response. However, simulations of this form of the model show that, in this case, either mean response times for equality responses at all levels of discriminability will be greater than times for the corresponding extreme responses, or they will be less (Vickers, 1979, p. 109). This prediction was not supported by Vickers' (1975) results, discussed above.

Vickers (1975, 1979) also proposed an extension of the accumulator model to account for equality judgments. According to this model, the observer stores observations of the stimulus difference in three accumulators, one of which corresponds to each of the possible responses. As in the two-category model, the magnitude of the evidence in favour of either extreme response is stored in the accumulator for that response. Evidence for an equality response is measured in terms of the sum of the totals t_g and t_l of evidence for the extreme responses minus the modulus of their difference. An equality response will be made when this value exceeds a critical value k_e, or, as described in an equation, when

$$(t_g + t_l) - |t_g - t_l| > k_e \tag{3}$$

Simulations of this model show that the predicted psychometric functions for the three responses take the same general form as those obtained by Kellogg (1931) and Vickers (1975) (Vickers, 1979). The simulations also show the commonly observed inverse relationship between times for extreme responses and stimulus difference. In general, in these simulations, times for equality responses are among the longest, taking values close to the times for extreme responses at small stimulus differences, but decreasing slightly as stimulus difference increases.

Other models, which are closely related to those described above, have been suggested to explain performance in three-category judgments. For example, Pike (1973) proposed a multiple observations version of the model of Swets and Green (1961) based on Signal Detection Theory. As in the Signal Detection Theory model, it is assumed that the observer adopts two cutoffs, x_g and x_l, on the distribution of stimulus difference. However, in this case the observer takes more than one observation, counting the number which fall above x_g, between x_g and x_l, and below x_l. The appropriate response ('greater than', 'equal to' or 'less than') is given when the total number of observations in favour of one response exceeds the criterion value for that response (r_g, r_e or r_l). Thus, the model involves the supposition of both sensory criteria (x_g and x_l) and response criteria (r_g, r_e and r_l), but since there is such a small amount of empirical evidence available in this area, it is doubtful whether this complexity is justified.

(d) Confidence in models for two-category judgment

One aspect of performance which was not originally included in these models is subjective confidence. Subsequently, two hypotheses have been suggested. The first suggestion was made by Audley (1960) and may apply to any of the models. As mentioned above, Audley (1960) proposed that confidence is an inverse function of the number of observations which occur before a response is made, and therefore, an inverse function of response time. This descriptive hypothesis can account for the direct variation of confidence and discriminability reported by Garrett (1922), Johnson (1939), Festinger (1943a) and Pierrel and Murray (1963). For example, when discriminability is low, many observations will be required to satisfy the criterion, so response time will be correspondingly long, and confidence will be low. The converse pattern will obtain when discriminability is high. In addition, the hypothesis can predict the effects of bias and caution on confidence. When the observer is biassed towards a particular response, it is assumed that the criterion value for this preferred response will be low. Therefore, the time to make the response will be short, and consequently confidence will be high. At the same time, the criterion value for the non-preferred response will be high, so these responses will be slow and less confident. On the other hand, when the observer is unbiassed but caution is high, it is assumed that the criteria for both responses will also be high. In consequence, both responses will be made with long response times and low confidence. Conversely, when caution is low, the criteria for both responses will be low, response times will be short, and confidence will be high.

Another mechanism for confidence was proposed by Vickers (1972) and may apply to a recruitment process or an accumulator model. Vickers (1972) suggested that the observer's confidence is a measure of the 'balance of evidence', or the difference between the stored amounts of evidence in favour of the two responses, when a decision is made. As with Audley's (1960) hypothesis, this formulation can account for the relationships between confidence and the independent variables of discriminability, bias and caution, as well as the dependent variables of accuracy and time.

For example, the mechanism predicts the direct relationship between confidence and discriminability. If discriminability is low, each observation will tend to yield only a small amount of evidence, while the proportions of observations which favour each response will be very similar. Therefore, at the time of decision, the totals of evidence in favour of each response will not differ very much, so confidence will be low. If discriminability is high, the majority of observations will yield a large amount of evidence in favour of the correct response. Therefore, at the time of decision, little evidence will have been collected in favour of the alternative response, so the difference between the totals of evidence will be large and confidence will be high.

If an observer is biassed towards a particular response, the criterion for this preferred response will be low. Therefore, when the observer makes this response, the maximum positive difference between the totals of evidence will be equal to the low criterion value for the response, so confidence will be correspondingly low. Alternatively, when the non-preferred response is made, the totals of evidence may differ by an amount equal to the high criterion value for the non-preferred response, and confidence will tend to be high. Vickers (1979) suggests that the negative values of confidence which might occur when errors are made in this situation could provide a basis for the observer's awareness of having responded in error.

Similarly, variations in confidence due to changes in caution can be predicted. When caution is high it is assumed that the criteria for both responses are high. As a consequence, the totals of evidence may differ by a greater amount than when caution is low and both criteria take correspondingly low values. Therefore, confidence will tend to be greater when caution is high.

The relationship between confidence and accuracy follows directly from the relationship between confidence and discriminability. Since confidence is higher for larger values of stimulus difference and these are discriminated more accurately, confidence will vary directly with accuracy. Similarly, the inverse relationship between confidence and response time, when nominal or average discriminability is held constant, can be explained in terms of trial to trial variations in discriminability. In this situation, a long response time indicates that a large number of observations was taken before the total of evidence in favour of one response reached criterion.

As a result, the evidence for the alternative response, which has not reached criterion, will necessarily be greater than if only few observations had been taken. Therefore, the difference between the totals of evidence will be small and confidence will be low.

(e) Confidence in models for three-category judgment

Audley's (1960) hypothesis that confidence is an inverse function of the number of observations taken before a response, may be applied in models for three-category judgment, in which case it would be expected that confidence would vary inversely with the times for each of the three responses. The predictions of this hypothesis, when only the extreme responses are considered, are outlined above. However, the inclusion of equality responses gives rise to the possibility of an equivalent level of bias towards both extreme responses when the observer is biassed away from equality responses, or *vice versa*. In this situation the hypothesis would predict that confidence should vary as an inverse function of the bias-induced changes in response time. Similarly, as in the case of two-category judgments, it would be expected that confidence will vary as an inverse function of changes in response time resulting from variations in caution.

Vickers' (1979) balance of evidence hypothesis has also been extended to include three-category tasks. By analogy with the two-category case, confidence is defined as the average difference between the total of evidence accumulated in favour of the given response and the totals of evidence for the other two possible responses. Therefore, if t_g and t_l are the totals of evidence in favour of 'greater than' and 'less than' responses, respectively, then the confidence in a 'greater than' response can be expressed as

$^1/_2$ [$|t_g - t_l| + (t_g - 2t_l)$]. Conversely, the confidence in a 'less than' response can be expressed as $^1/_2$ [$|t_g - t_l| + (t_l - 2t_g)$]. Accordingly, the confidence in an equality response will be the average of these two values, or, $^1/_2$ [$(t_g + t_l) - 2|t_g - t_l|$]. Simulations of this model show that confidence in each of the extreme responses tends to decrease as stimulus difference decreases while confidence in equality responses tends to remain fairly constant (Vickers, 1979). Once more it is predicted that confidence will vary with changes in bias and caution. Simulations of the process show that, for both extreme responses and equality responses, confidence is greater for a non-preferred response for which the criterion value is relatively high (Vickers, 1979). Similarly, when caution is high so that the criteria for all responses are high, it is predicted that confidence will be high.

C. CRITERION REGULATION MECHANISMS

A number of different mechanisms has been suggested for the regulation of criteria in the above decision models. Although they have usually been applied to two-category judgments, in most cases they can also be applied to the three-category case.

1. The ideal observer hypothesis

The actual performance of the observer in a discrimination task may be compared with a normative model embodied in the so called 'ideal observer' hypothesis (Corso, 1967; Egan, 1976). This hypothesis is most commonly applied to Signal Detection Theory but it can also be applied to multiple observation models, and can regulate criterion changes in both two-category and three-category tasks. In terms of Signal Detection Theory, the extension to the three-category case becomes clearer if both extreme stimuli are classed together as occurrences of the signal, and the equal stimuli are regarded as the presentation of noise alone, so that the task is considered as a same-different judgment. In this case, as for two-category tasks, the sensory effect of the stimulus difference may be represented by a normal distribution on which the observer adopts a single cutoff.

According to the ideal observer hypothesis, the observer chooses a criterion value, β, rationally, in such a way that he maximises the expected value of his decisions. In order to do this, when choosing β, the observer must take into account the a priori stimulus probabilities and the prescribed costs and payoffs associated with making each response correctly or incorrectly. In the detection situation, the optimal value of β is defined as

$$\beta i = \frac{p(n)}{p(s)} \cdot \frac{(V_{Nn} + C_{Sn})}{(V_{Ss} + C_{Ns})} \qquad (4)$$

where p(n) and p(s) are the a priori probabilities of noise alone and of signal plus noise and V_{Ss} and V_{Nn} are the values associated with hits and correct rejections, and C_{Sn} and C_{Ns} are the costs of false alarms and misses, respectively. In this section, notation in terms of signal (s) and noise (n) will be used for the sake of simplicity. According to the convention followed in this thesis, a signal may be regarded as corresponding to a stimulus V>S, while noise corresponds to a stimulus V<S.

The criterion, β, may be expressed in terms of a likelihood ratio (LR), which is defined as the likelihood that a particular observation, x, arose when a signal was present, or, in equation form

$$LR = \frac{p(x|s)}{p(x|n)} \qquad (5)$$

where $p(x|s)$ is the conditional probability of observation x given that a signal was present and $p(x|n)$ is the conditional probability of observation x given that there was no signal. When the two stimulus events are equiprobable, the LR gives the odds in favour of a signal occurring. When the stimulus events are not equiprobable, the actual probabilities must be taken into account in the determination of the LR. Similarly, asymmetrical costs and payoffs will influence the odds that a signal response will be made. Thus, Egan (1976, pp. 11-18) shows that the choice of a criterion, β, may be interpreted as the selection of a desired LR or minimum odds that a signal occurred, and that a decision is made on the basis of whether the obtained LR is greater or less than the desired LR.

The formula for β (see equation (4)) implies that, if the payoffs are symmetric and constant, then a change in the a priori probabilities of the stimuli will produce a change in β. For example, if $p(s) = 0.2$, so that $p(n) = 0.8$, the adopted criterion will take the value, $\beta = p(n)/p(s) = 0.8/0.2 = 4.0$. However, if p(s) increases so that $p(s) = p(n) = 0.5$, according to the hypothesis, the criterion will decrease to take the value $\beta = 0.5/0.5 = 1.0$. Conversely, when the probability of a signal decreases, β will increase. This increase in the stringency of the criterion as the stimulus probability decreases has often been proposed as an explanation for the decrease in the percentage of hits which commonly occurs in vigilance tasks as the experimental session progresses (e.g. Loeb & Binford, 1964, 1968; Colquhoun & Baddeley, 1964, 1967; Broadbent & Gregory, 1965; Jerison, Pickett & Stenson, 1965; Mackworth, 1965; Taylor, 1965; Binford & Loeb, 1966; Colquhoun, 1966, 1967, 1969; Levine, 1966; Baddeley & Colquhoun, 1969; Williges, 1969).

Since the aim of choosing a criterion according to the ideal observer hypothesis is to maximise the expected value of each decision, the effects of a change in the value of p(s) can also be examined in terms of the resulting change in expected value which is defined as

$$E(V) = p(S|s)\ p(s)\ V_{Ss} + p(S|n)\ p(n) C_{Sn} + p(N|n)\ p(n)\ V_{Nn} + p(N|s)\ p(s)\ C_{Ns} \qquad (6)$$

For example, when p(s) = 0.2 and therefore p(n) = 0.8, $\beta = 4.0$, as seen above. If it is assumed that d' = 1.0 then, from Freeman (1964), p(S|s) = 0.2 and p(S|n) = 0.03. Therefore, if the payoff matrix is symmetric with unitary values so that $V_{Ss} = V_{Nn} = 1.0$ and $C_{Sn} = C_{Ns} = -1.0$, equation (6) becomes

$$E(V) = (0.2)\ (0.2)\ (1.0) + (0.03)\ (0.8)\ (-1.0) + (0.8)\ (0.8)\ (1.0) + (0.97)\ (0.2)\ (-1.0) = 0.46$$

If the payoff matrix remains constant and the <u>a priori</u> probabilities change so that p(s) = p(n) = 0.5 then, from Freeman (1964), p(S|s) = 0.69 and p(S|n) = 0.3. In this case,

$$E(V) = (0.69)\ (0.5)\ (1.0) + (0.3)\ (0.5)\ (-1.0) + (0.7)\ (0.5)\ (1.0) + (0.32)\ (0.5)\ (-1.0) = 0.39$$

When the probability of signal plus noise increases further so that p(s) = 0.8 and p(n) = 0.2, then p(S|s) = 0.97 and p(S|n) = 0.8 (Freeman, 1964). In this case,

$$E(V) = (0.97)\ (0.8)\ (1.0) + (0.8)\ (0.2)\ (-1.0) + (0.2)\ (0.2)\ (1.0) + (0.03)\ (0.8)\ (-1.0) = 0.63$$

Therefore, it can be seen that the expected value of a decision decreases as the probability of a signal occurring increases from less than, to equal to the probability of noise alone. When p(s) increases still further to take a value greater than p(n), the expected value increases.

In addition, according to equation (4), if the stimulus probabilities are constant, the criterion value may be manipulated by variations in the payoff matrix. For example, if $p(s) = p(n) = 0.5$, an observer may be asked to maximise his earnings when $V_{Ss} = V_{Nn} = 2.0$ and $C_{Sn} = C_{Ns} = -1.0$. In this case, according to equation (4)

$$\beta = \frac{0.5}{0.5}\ \frac{(2.0 - 1.0)}{(2.0 - 1.0)} = 1.0$$

If the value of hits is doubled so that $V_{Ss} = 4.0$ while all the other costs and payoffs remain the same then

$$\beta = \frac{0.5}{0.5}\ \frac{(2.0 - 1.0)}{(4.0 - 1.0)} = 0.3$$

The lower criterion increases the proportion of hits that will be made so that the overall earnings of the observer will increase. However, many studies have failed to find any variation in the empirical values of β when payoff matrices are varied (e.g. Lieblich & Lieblich, 1969a,b; Williges, 1971; Guralnick, 1972). Swets (1977) noted that the variability in the effects of payoffs on performance was not surprising when one considered that maximising earnings was often dependent on the observer making errors. This would appear to conflict with the preference of most observers for accurate performance. In addition, the matrix itself may have little meaning in the situation where there is a long time delay between the allocation of points during the experimental session and their conversion to fractions of a cent at the end of the day. Moreover, Swets (1977) pointed out that, as the experimental situation becomes closer to the experiences of the observer in day to day living, the subjective values and costs of the situation to the observer, based on his previous experience, may outweigh any experimental manipulation.

An important property of the ideal observer hypothesis is that it assumes that the observer's criterion is regulated in response to changes in a priori probabilities. As in the ideal observer hypothesis, a priori probabilities may determine the criterion value in the random walk model. In the random walk model, the observer's estimate of a priori probabilities is represented by the distance between the starting point and each boundary. It follows that, changes in the criterion value as a result of changes in a priori probabilities, will correspond to changes in the relative distance between the starting point and the two boundaries. This may be achieved by movement of the boundaries or of the starting point, or both. However, since Laming (1968, p. 33) and Link and Heath (1975, p. 91) have shown that these are equivalent changes, following Ascher (1974) and Link (1978a) it will be assumed that variations in a priori probabilities give rise to a change in the starting point. For example, an increase in the probability of the stimulus V>S (signal) will produce a movement of the starting point closer to the boundary corresponding to responses 'V>S'. In figure 6 this is represented by a move from SP_1 to SP_3.

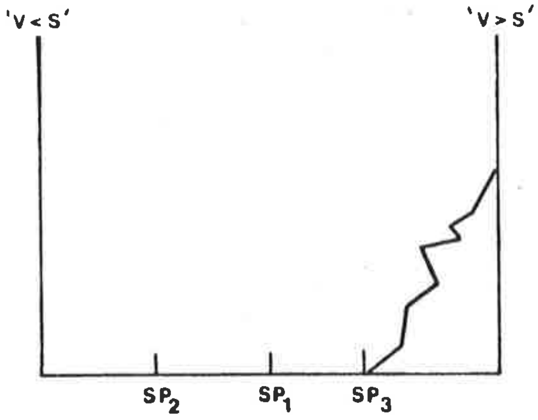


FIGURE 6. Schematic representation of the random walk model showing three possible positions of the starting point, favouring neither response (SP_1), favouring responses 'V<S' (SP_2) and favouring responses 'V>S' (SP_3). The diagram shows the form of a hypothetical random walk made by an observer who is biassed towards responses 'V>S' when a stimulus V>S is presented.

Similarly, changes in the criterion value of the ideal observer hypothesis due to variations in costs and payoffs can also be accounted for in terms of a movement in the starting point. In this process, the distance between the starting point and the appropriate boundary will determine the number of observations required before a response is made and therefore, the time taken to make it. Since this distance, the LR, represents an estimate of the probability that the response is correct, it appears likely that confidence will vary as a direct function of this distance, or the size of the LR.

In the accumulator model, the observer's estimate of a priori probability is also represented by the relative values of the criteria for the two responses. Therefore, variations in the criterion as a result of changes in a priori probability, will produce changes in the criterion values. Criterion values will also vary with changes in costs and payoffs.

2. The response stabilisation hypothesis

Another mechanism which was suggested to regulate criterion values in Signal Detection Theory is the response stabilisation hypothesis of Vickers, Leary and Barnes (1977). According to Vickers et al. (1977), the observer adjusts his criterion value in order to minimise any discrepancy between the local probability of a signal response (L, estimated over a limited number of recent trials) and the cumulatively determined probability of making signal responses (C, estimated over the entire experimental session including demonstration and practice trials). When the discrepancy between these values exceeds some critical amount, the criterion is adjusted. If L exceeds C, the criterion is adjusted upwards and if L is less than C the criterion is adjusted downwards.

An equivalent mechanism was outlined by Indlin (1976) who used the analogy of the observer acting as a system with feedback. In this system, the reference signal is the proportion of a given type of response made over the entire experimental session (corresponding to C) and the feedback signal is the proportion of these responses made within a set time interval prior to the current response (corresponding to L). A mismatch between the reference signal and the feedback will give rise to a change in the criterion in exactly the same way as described above for the response stabilisation hypothesis.

For example, an observer may perform a block of practice trials in which stimuli are equiprobable and then change to experimental trials in which the probability of a stimulus V>S is 0.8. In this case, in the practice block, L and C will both tend to a value of 0.5, so there will be no change in the criterion. However, when the experimental trials begin, the local probability of a response 'V>S' will increase quickly to 0.8, while the cumulative probability of this response will increase more slowly. Therefore, L will be greater than C and the criterion will be adjusted upward. At the same time, the value of L for the alternative response 'V<S' will have dropped to 0.2 while the corresponding value of C will be decreasing more slowly. Consequently, L<C and the criterion for the response 'V<S' will be adjusted downward.

According to Vickers et al. (1977), in terms of Signal Detection Theory, if L>C for the response 'V>S', the cutoff will be adjusted, for example, from x_{c_1} to x_{c_2} in figure 7, so that observations will be less likely to fall on the side of the cutoff corresponding to responses 'V>S'. Conversely, if L<C for the response 'V>S', the cutoff will move to a position

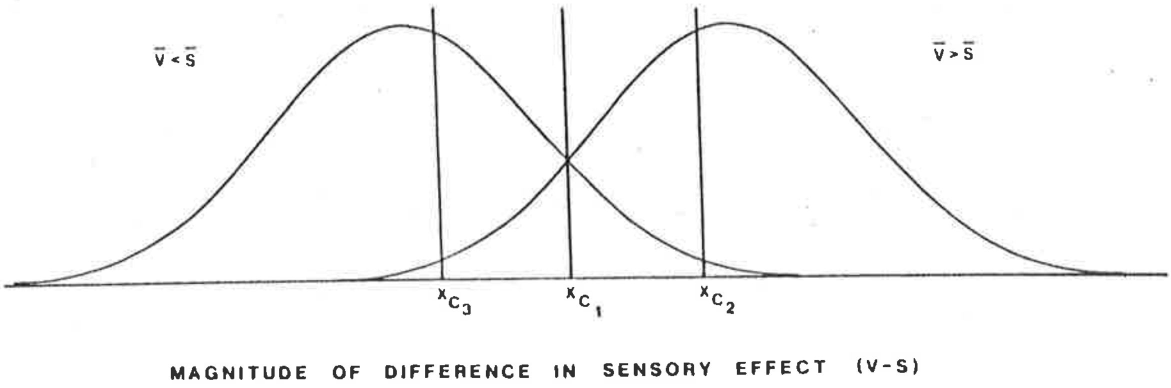


FIGURE 7. Probability density distributions of stimulus difference (V-S) appropriate to Signal Detection Theory for two-category tasks, showing three possible positions of the cutoff, favouring neither response (x_{c_1}), favouring responses 'V<S'(x_{c_2}), and favouring responses 'V>S' (x_{c_3}).

where observations are more likely to favour this response (e.g. position x_{c_3} in figure 7).

Alternatively, the response stabilisation hypothesis can apply to a random walk type of process. As the observer's estimate of a priori probability is represented by the distance between the starting point and the boundaries, this mechanism of criterion regulation can be assumed to determine the position of the starting point. In this case, when L>C for the response 'V>S', the starting point will move further from the boundary for this response, for example from position SP_1 to SP_2 in figure 6. When L<C for the response 'V>S', the starting point will move closer to the boundary for this response (e.g. position SP_3 in figure 6).

When the response stabilisation hypothesis is applied to an accumulator process, variations in the relative values of L and C will also produce changes in the criterion values. For example, when L>C for the response 'V>S', the criterion for this response will increase relative to the criterion for the alternative response. Conversely, when L<C for the response 'V>S', there will be a relative decrease in the criterion for this response.

3. The 'adaptation-level' hypothesis

A third possible mechanism to regulate criteria within the framework of Signal Detection Theory was outlined by Vickers and Leary (1983). This mechanism does not presuppose that the observer has such detailed knowledge of stimulus properties such as probabilities, but regulates criteria on the basis of information from the previous stimuli in the experiment. According to this formulation, when the payoff matrix is symmetric, and the observer is not informed of the changes in signal probability, he adopts a cutoff value which is equal to the mean of the sensory intensities of all observations experienced up to that point in the experiment.

For example, if the probability of a stimulus V>S changes from 0.5 in practice trials to 0.8 in experimental trials, the number of observations which favour the corresponding response will also increase. According to the adaptation-level hypothesis, this will result in an increase in the criterion for this response. This mechanism can be applied to Signal Detection Theory and multiple observation models.

According to Vickers and Leary (1983), when the adaptation-level hypothesis is applied to Signal Detection Theory, in the situation where most observations favour the response 'V>S', the cutoff will move to a position where observations will be more likely to favour the alternative response. For example, the cutoff would move from x_{c_1} to x_{c_2} in figure 7.

In the random walk model, Link and Heath (1975) proposed that each observation of successively presented stimuli is compared with a subjective referent in order to determine the direction and magnitude of the evidence it provides. As Link and Heath (1975) noted "the referent must depend upon prior exposure to the two stimuli, and, furthermore, may be shaped or emphasized by the number of previous exposures to each stimulus" (p. 78). Therefore, in this model it may be assumed that the adaptation level is directly represented by the subjective referent. Correspondingly, when stimuli are presented simultaneously, it appears that any categorisation of observations must take place with reference to a neutral or indifference point, similar to the subjective referent, which separates those observations which favour one response from those which favour the alternative. It appears likely that, as with the subjective referent, the value of the indifference point will be influenced by previous experience of the stimulus sequence and therefore, that its value will be directly related to the adaptation level. For example, if the probability of a stimulus V>S increases from 0.2 to 0.8, the indifference point will increase. As a consequence the discriminability of stimuli V>S is effectively reduced.

Similarly, in the accumulator model, it can be assumed that the adaptation-level hypothesis determines the value of an indifference point which divides those observations which favour a response 'V>S' from those which favour a response 'V<S'. As for the random walk model, an increase in the probability of stimuli V>S will produce an increase in the value of the indifference point or, effectively, a reduction in the discriminability of these stimuli.

4. The target confidence mechanism

Another criterion regulation mechanism was developed by Vickers (1979) as part of the accumulator model. Vickers (1979) suggested that the observer uses a measure of the confidence which he feels in his judgments to regulate criterion values in such a way that he attempts to maintain a certain 'target' level of confidence on average in all of his judgments.

Vickers (1979) proposed that, for each response, there is a pair of control accumulators. After a given response, the difference between the actual confidence at the time of decision and the target level of confidence, k, is stored in the control accumulators. Overconfidence, when actual confidence is greater than k, is stored in one accumulator, while underconfidence, when actual confidence is less than k, is stored in the other. If either the total of overconfidence or the total of underconfidence reaches the criterion value of the corresponding control accumulator, the criterion for the response in the primary decision process is adjusted upwards or downwards, respectively. The amount by which the criterion changes will be proportional to the difference between the totals of overconfidence and underconfidence stored in the control accumulators at the moment when one of them reaches criterion. For example, when an observer adopts a high level of caution it would be assumed that his target levels of confidence for both responses would be high. Initially it could happen, for example, that an observer's confidence in both responses is lower than the target level. Therefore, underconfidence would be collected in the control processes for both responses. As a consequence, there would be an upward adjustment of the criteria for both responses by an amount proportional to the net underconfidence in each control process.

In Signal Detection Theory, if response time is determined by the inclusion of latency functions (e.g. Gescheider et al., 1968, 1969; Pike, 1973) and Audley's (1960) hypothesis is the basis for confidence, it is assumed that confidence is directly related to the distance between an observation and the cutoff, so that an increase in confidence in one response can only be achieved at the expense of a reduction in confidence in the alternative response. Therefore, it is not possible to increase or decrease confidence in both responses simultaneously as would be expected when caution increased or decreased. Similarly, it is not possible to maintain the same target level of confidence in both responses. Consequently, the notion of maintaining a target level of confidence cannot be applied to Signal Detection Theory as it stands.

The mechanism based on target confidence can be applied to a random walk model. As seen above, the criterion in the random walk model can be interpreted as the attainment of a fixed LR. In this case, since the LR is an estimate of the probability that the response is correct, it appears that confidence should be directly related to the value of the LR. As a consequence, all responses should be made with the same confidence. However, as this is not the case in empirical studies, it appears that the LR is not fixed but that it may vary randomly from trial to trial. In other words, the relationship between the starting point and the boundaries varies at random from trial to trial. In this case, it has been suggested that confidence varies as a direct function of the relative values of the starting point and the boundaries, but is also influenced by the drift on a given trial (Ascher, 1974; Link, 1978a). In equation form

$$C = K\,[\Theta(A-S)] \qquad (7)$$

where K is a scaling constant, the response boundaries have values of A and -A, Θ is the drift on a given trial and S is the starting point with $-A<S<A$. However, Pike (1968, p. 173) has shown that, when the criterion varies in this way, errors tend to be made with shorter response times than correct responses because the longer responses associated with the higher criterion values will have a greater probability of being correct. In contrast, in most discrimination experiments, errors are slower than correct responses. Therefore, it appears that, if the observer is using an LR decision rule, the criterion remains relatively stable. For example, Sanders and Ter Linden (1967) have suggested that the criterion does not vary randomly but that it diminishes as a regular function of time. In this case, as time elapses the distance between the starting point and either boundary decreases (see figure 8). If, as suggested

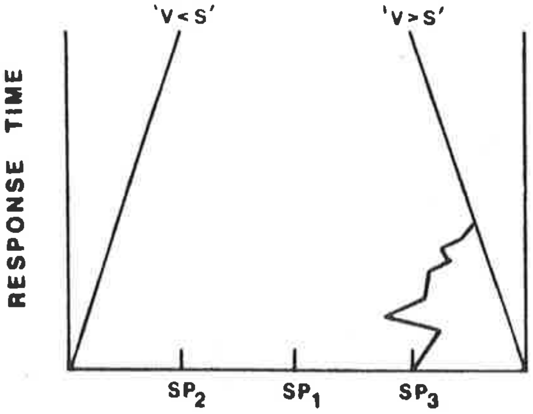


FIGURE 8. Schematic representation of the random walk model with diminishing criterion, showing three possible positions of the starting point, favouring neither response (SP_1), favouring responses 'V<S'(SP_2) and favouring responses 'V>S'(SP_3). The diagram shows the form of a hypothetical random walk made by an observer who is biassed towards responses 'V>S' when a stimulus V>S is presented.

above, confidence is a direct function of the LR, then, as time elapses and the LR decreases, confidence will also decrease. In other words, in this process confidence will vary inversely with response time, as proposed by Audley (1960). This form of the model appears to be more consistent with empirical results. If the target confidence mechanism is applied to a random walk model with a diminishing criterion, the process operates in a way which is roughly similar to an accumulator model with the target confidence mechanism. In order to maintain the target level of confidence the position of the starting point relative to the boundaries will change. For example, if confidence for a response 'V>S' is consistently lower than target confidence then the distance between the starting point and the boundary for this response will increase. As before, it is convenient to assume that the starting point moves rather than the boundaries. Therefore, this corresponds to a movement of the starting point from SP_1 to SP_2 in figure 8. Similarly, if confidence in the response 'V>S' is consistently higher than target confidence, the starting point will move closer to the boundary for this response or for example, to position SP_3 in figure 8.

On the present statement of the models with variable criteria as described above, it appears that the manipulation of instructions, payoffs or stimulus probabilities should produce equivalent changes in performance. In each of the models, these manipulations produce changes in performance which are mediated by changes in criterion values. However, it is apparent that the extent to which each model predicts the exact change in performance to result from any of these manipulations depends on the precise form of the mechanism which regulates the criterion values in the model. For example, the ideal observer hypothesis of Signal Detection Theory predicts that the effects of instructions, payoffs and stimulus probabilities will be equivalent, since each of these possible manipulations is always taken into account in the determination of the criterion value according to this mechanism. In contrast, in the other three mechanisms for criterion regulation, the contribution of the stimulus characteristics such as discriminability and probability is quite explicit while the role of instructions and payoffs is not as well defined. As a consequence, although it seems obvious that criteria must be susceptible to the effects of payoffs and instructions, it is not clear that these mechanisms predict that the effects of these manipulations will be equivalent. In addition, there is little empirical evidence to help determine the relative effects of these manipulations. For example, comparatively few studies have compared the effects of stimulus probability and instructions (e.g. Williges, 1969) or of stimulus probability and payoffs (e.g. Williges, 1971).

As shown by the review of the literature, there is no empirical evidence in many of the areas which would appear to be relevant to any assessment of the effects of criterion change. For example, although three-category judgments appear to be sensitive to the effects of a change in attitude such as may be induced by biassing instructions, this area has received little attention. Similarly, the effects of bias in two-category judgments have not been examined in detail, and confidence measures have never been taken in this situation. In addition, comparisons between the different forms of experimental manipulations have rarely been undertaken.

At the same time, even less attention has been paid to the relationships between dependent variables. Indeed, the only relationship which has been studied is that between accuracy and time, and this relationship itself has been important in pointing out the inadequacies of taking a single response measure. As seen above, the nature of the speed-accuracy tradeoff has shown that the interpretation of either speed measures or accuracy measures on their own would be somewhat ambiguous. Even when two dependent variables are recorded, the situation may be ambiguous, as one may change when the other does not. For example, in the study of Wilding (1974), measures of accuracy and response time were recorded at different levels of caution. Although there was no change in accuracy as caution was increased (implying that criterion values had not been influenced by instructions), there was a systematic increase in response time with increasing caution (which would appear to be consistent with an increase in the criterion values). For this reason it appears to be preferable, wherever possible, to study the effects of all three dependent variables.

Initially the present thesis was aimed at an examination of criterion regulation when response bias was induced in a three-category task, since this form of task appeared to be particularly susceptible to the effects of changes in attitude. In addition, it appeared that the measurement of all three dependent variables of accuracy, time and confidence would produce less equivocal results than previous studies in which only one or two response measures were taken. Two experiments which provide data on this problem are presented in Chapter 3. However, as the thesis progressed, other areas emerged as central to the problem of criterion control and therefore were investigated. In particular, it appeared that a comparison between the effects of the independent variables would provide important evidence for an evaluation of the criterion regulation mechanisms. Therefore, in Chapters 4 and 5, three experiments on the biassing effects of a priori probability in a two-category task are reported, while the two experiments in Chapter 6 examine instruction-induced bias in the same task. It was also apparent that one method of biassing responding, that of changing the relative difficulty of the possible responses, had been largely overlooked. This form of experimental manipulation is examined in the two experiments described in Chapter 7. In addition, there was a major inconsistency in the findings of previous studies on the effects of manipulating caution. Data on this problem are presented in Chapter 8. Finally, in Chapter 9, the empirical findings from the manipulation of stimulus probabilities, discriminability and instructions are compared and examined in terms of the decision processes.

CHAPTER 3

A. EXPERIMENT 1

1. Introduction

In Chapter 1 it was seen that many experimental studies have examined the choice between two extreme responses in which the variable stimulus is judged as either greater than or less than the standard. Less attention has been paid to tasks involving three responses in which the third, intermediate response, indicates the perception of equality, or no difference between the stimuli. This comparative neglect is surprising, since the judgment of equality seems intuitively to complement the judgment of difference.

In particular, in Chapter 1 it was noted that few studies using three-category tasks have examined the effects of bias, even though equality judgments appear to be highly responsive to changes in attitude. For example, in early studies in which there was no explicit attempt to bias responding, results implied that observers had a natural bias away from equality responses. Equality responses tended to be less probable, slower, and less confident than extreme responses (e.g. Hayden, 1906; Angell, 1907). However, Vickers (1975) has shown that these trends can be changed when bias is induced by variations in *a priori* probabilities. As seen above, Vickers (1975) found that when the *a priori* probability of equal stimuli changed from low to high there was a corresponding increase in the probability of accurate equality responses. In addition, the times for correct equality responses changed from longer to shorter than times for correct extreme responses at the smallest stimulus difference. Equivalent results were obtained by Fernberger (1914a,b, 1931) when instructions emphasising the validity of equality responses were used. However, both Fernberger (1914a,b, 1931) and Vickers (1975) recorded only accuracy and response time measures. No published data are known on the effects of bias on confidence.

Similarly, the effects of variations in discriminability on accuracy and response time in three-category tasks have seldom been studied (Kellogg, 1931; Vickers, 1975). As noted in Chapter 1, Kellogg (1931) and Vickers (1975) found that the probability of making an extreme response decreased as the stimulus difference decreased, while the probability of an equality response increased as the stimulus difference approached zero. Both Kellogg (1931) and Vickers (1975) noted an inverse relationship between response time and discriminability for correct extreme responses, while correct equality responses took the same time or were slower than correct responses to the smallest stimulus difference. However, no data are known in which the relationship between discriminability and confidence in three-category tasks is considered.

Typically, experiments on the effects of variations in bias have employed either a between-subjects design or a within-subjects design, while discriminability is commonly varied within subjects. Both of these forms of experimental design have advantages and each involves some problems. This is well illustrated when the manipulation of bias is considered. When bias is manipulated in a between-subjects design, it is possible to examine a large number of bias conditions using short experimental sessions, because each subject performs under only one bias condition. As a result, fatigue and boredom effects can be minimised. However, in a between-subjects design the effects of individual differences will be included in any comparison of bias conditions. Therefore, the variance due to individual differences may mask the effects of bias. On the other hand, when a within-subjects design is used, it is only practicable to examine a restricted range of bias conditions within any one experimental session. Otherwise, within any one condition, insufficient measures would be obtained to be reliable. Even when the number of bias conditions is restricted, this form of design tends to result in lengthy experimental sessions in which carry-over effects may become important (Winer, 1972). One alternative is to administer the different bias conditions to each observer in separate experimental sessions. However, this too is likely to involve problems with motivation, and the drop-out rate of observers may be high. The major advantage of a within-subjects design is that the effects of bias are not confounded with the variance arising from individual differences.

As an initial investigation of the effects of bias on accuracy, time and confidence, a three-category task was used, since equality responses appear to be very responsive to changes in bias. Variations in bias were induced by changes in a priori probability following Vickers (1975). At the same time, since variations in discriminability did not appear to mask the effects of bias in Vickers' (1975) study, it was decided to examine those effects also. Therefore, data on confidence and discriminability could be obtained to complement the findings of Vickers (1975) and Kellogg (1931) on accuracy and time. In this first study, a between-subjects experimental design was used so that a broad range of bias conditions could be examined.

2. Method

(a) Stimuli

The stimuli were identical to those used by Vickers (1975), consisting of two horizontally adjacent square arrays each comprising 80 horizontal line segments arranged in 8 colinear rows and 10 columns. In each column, the left endpoints of each segment were colinear (see figure 9). The lengths of the line segments were randomly determined from a truncated normal distribution with a constant standard deviation of 5 display units (approximately 0.075 cm). The standard array with mean segment length of 15 display units (approximately 0.225 cm) was positioned at random to the left or the right of

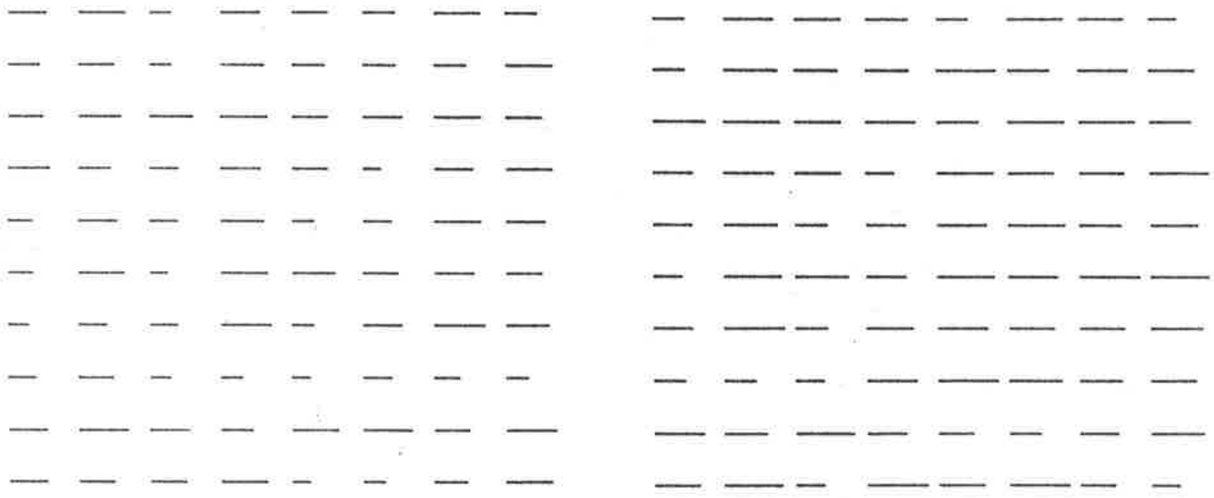


FIGURE 9. An example of the type of stimulus used in Experiment 1. In this drawing the mean segment length in the right array is 21 display units and the mean segment length in the left array is 15 display units.

a variable array in which the mean segment length was 15, 17, 19, 21, 23 or 25 units. Therefore, there were six levels of discriminability or stimulus difference, one in which the mean segment lengths of the standard and variable arrays were equal and five in which they were different. These six levels have been labelled 0, 1, 2, 3, 4 and 5, respectively.

(b) Apparatus

Stimuli were generated by a PDP 8/L computer and plotted in refresh mode on a Tektronix 611 display. Immediately in front of the observer was a table of response keys comprising a centrally located triangular marker positioned equidistant from each of 3 telegraph keys placed about 6 cm further away from the observer. Adjacent telegraph keys were 4 cm apart, centre to centre.

Responses were made by moving the right index finger from the apex of the triangular marker to the appropriate key. Depression of any of the keys by about 3 mm was necessary to make contact which was accompanied by a discernible 'click'. Presentation of a stimulus was terminated by the observer's response, or after 5 sec, whichever occurred first. For each response, the time between the presentation of the stimulus and the key press was measured and recorded by the computer.

Observers gave verbal ratings of their confidence after each response. These ratings were recorded on a cassette tape recorder which was connected to a microphone positioned around the observer's neck.

(c) Observers

The 36 observers were students enrolled in the first year psychology course at the University of Adelaide whose participation was credited towards a course requirement. They ranged in age from 17 to 24 years and all declared themselves to be right-handed. All observers were naive with respect to the aims of the experiment.

(d) Design

Four groups each of 9 observers were formed. Each group performed in a different experimental condition. Observers were tested individually for two sessions both of which were approximately 45 minutes in length.

The four conditions were distinguished by the proportion of equal stimuli in the stimulus sequence. The proportion of equal stimuli remained constant within all blocks of trials in any given condition. In Conditions 1 to 4 the proportions of equal stimuli were 0.099, 0.455, 0.546 and 0.818, respectively. (These proportions result from the use of five levels of discriminability for each of right and left stimuli and an equal stimulus. Therefore, any balanced stimulus sequence must comprise multiples of eleven stimuli.)

In each session, stimuli were presented in three blocks, one block of 110 stimuli preceding two blocks each of 225 stimuli. Within all blocks of trials, the occurrence of the longer mean segment length in the right or the left array was equiprobable and followed a random order. In addition, each of the five levels of discriminability for extreme stimuli appeared with equal probability, and the sequence was random. Blocks were separated by short rest periods in which the observer was able to ask questions to clarify any procedural details.

The first experimental session and the first block of trials in the second session were treated as practice trials. In addition, the first five trials in each block were treated as warm-up trials and discarded. Only the data from the final two blocks of the second experimental session were considered in the statistical analysis. In this way, only well practised performance in which little further improvement could be expected was examined.

(e) Procedure

The experiment was performed in a darkened room. Approximately five minutes were allowed for adaptation to these conditions before commencing. The observer, seated at the table of response keys, was about one metre directly in front of the display screen. He was instructed that, on each trial, he should decide whether the mean segment length in one set of lines was longer than the other, or whether he considered the two sets to be equal in length. To indicate a decision, he was asked to press one of the three response keys, the left key if the left array were judged longer, the right key if the right array were judged longer, or the central key if the arrays were judged to be of equal average length. The observer was required to respond using the index finger of his right hand, replacing his finger on the apex of the triangular marker at the completion of each response. Observers were instructed to respond as quickly and accurately as possible.

In addition, the observer was asked to rate his confidence in the accuracy of each response. It was explained that, after each key press, he should indicate his confidence in that response by stating a number from 1 to 10, where 1 indicated the lowest possible confidence rating and 10 the highest possible rating. This ten point scale was chosen because it allowed for the expression of a large range of ratings, while being easy for the observer to use. The experimenter stressed that the confidence rating should be considered only after a key had been pressed, and that there was always a 2 sec interval in which this judgment could be made following the termination of each stimulus presentation. As a result, it was hoped to avoid any inappropriate delay in responding due to intervening decisions about confidence ratings.

3. Results

Although strictly, the rating of confidence is only an ordinal measurement, parametric significance tests, namely analyses of variance, were used in its examination in all of the following experiments. This was necessary due to the lack of an appropriate alternative. However, consideration of the robustness demonstrated by analysis of variance overall, and in particular, when the assumptions concerning the scale of measurement of the dependent variable are violated (Anderson, 1961; Glass, Peckham & Sanders, 1972) showed that this was unlikely to result in any misleading conclusions.

In addition, in all experiments percentage scores were used to measure the accuracy of responding. Since the percentage distribution is truncated, a normalising transformation (2 arcsin $\sqrt{p/100}$) was applied to the accuracy measures before any parametric tests of significance were undertaken (Winer, 1972; Cohen & Cohen, 1975). In all cases, post hoc comparisons between means were performed according to the Newman-Keuls procedure with a Bonferroni adjustment to give a family-wise error rate of $\alpha = .05$ (Winer, 1972). Tabulations of the complete results of all analyses of variance are presented in the appendices which are numbered to correspond to the experiments.

In this experiment, analyses of variance were carried out according to a crossed factorial design with Condition (a priori probability of equal stimuli, 4 levels as defined in section A.2(d)) and Discriminability (stimulus difference, 6 levels as defined in section A.2(a)) as factors. There were repeated measures on Discriminability. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses.

(a) Accuracy

The analysis of accuracy data showed significant main effects of Condition ($F(3,32) = 5.46$, $MSerror = 0.40$, $p < .001$) and Discriminability ($F(5,160) = 114.28$, $MSerror = 0.10$, $p < .001$) and a significant Condition x Discriminability interaction ($F(15,160) = 2.70$, $MSerror = 0.10$, $p < .01$).

The Condition main effect arose because accuracy was significantly higher in Conditions 1 and 3 than in Conditions 2 and 4. The percentages of stimuli for which responses were correct in Conditions 1 to 4 were 75.11%, 65.79%, 76.56% and 63.14%, respectively.

The Discriminability main effect reached significance because accuracy increased with stimulus difference, but at a decreasing rate. There was no difference in accuracy at Discriminability levels 0 and 1 and levels 4 and 5, while, at all other times, accuracy differed significantly between different values of stimulus difference. Accuracy rates for Discriminability levels 0 to 5 were 45.73%, 46.92%, 65.75%, 79.88%, 92.57% and 90.06%, respectively.

From table 2 it can be seen that the Condition x Discriminability effect was significant because the pattern of data was the same in Conditions 1, 2 and 3, where the probability of equal stimuli was lower than, or approximately equal to, the probability of extreme stimuli, while these three conditions contrasted with Condition 4, where equal stimuli were highly probable. In Conditions 1, 2 and 3, accuracy was lowest for responses to equal stimuli and increased with the size of the stimulus difference. In Condition 4 there was also a direct relationship between accuracy and stimulus difference for extreme responses, but at all levels of discriminability the accuracy of responses to extreme stimuli tended to be lower than in the other conditions.

	STIMULUS DIFFERENCE					
CONDITION	0	1	2	3	4	5
1	40.39	50.92	73.21	89.06	97.91	99.17
2	43.50	45.38	63.56	74.45	82.34	85.48
3	46.87	54.28	72.43	92.38	93.69	99.72
4	52.16	37.09	53.81	63.61	96.33	75.86

TABLE 2. Percentage of stimuli for which responses were correct at each of the six values of stimulus difference in each condition. The percentages were averaged over the nine observers in each condition.

In addition, in Condition 4, accuracy for responses to equal stimuli was higher than for responses to the smallest stimulus difference, and equality responses were made with greater accuracy than in any other condition. This change in the relative accuracy of equality and extreme responses from Conditions 1, 2 and 3 to Condition 4 can be seen in figure 10 where Conditions 1 and 4 are contrasted.

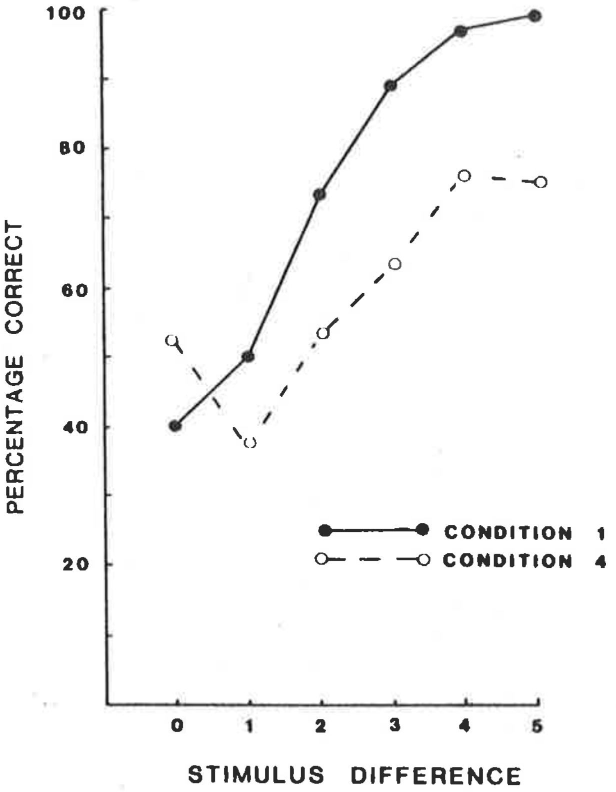


FIGURE 10. The percentage of stimuli for which responses were correct at each of the six values of stimulus difference in Condition 1 (solid line) and Condition 4 (broken line). The percentages were taken over the nine observers in each condition.

Figure 11 shows the psychometric functions for right, left and equality responses in each condition. It can be seen that, in each condition, the functions for the two extreme responses are ogival, while the function for equality responses is bell-shaped. From the figure there appears to be some suggestion of systematic change in the shape and position of these curves as a priori probability changes.

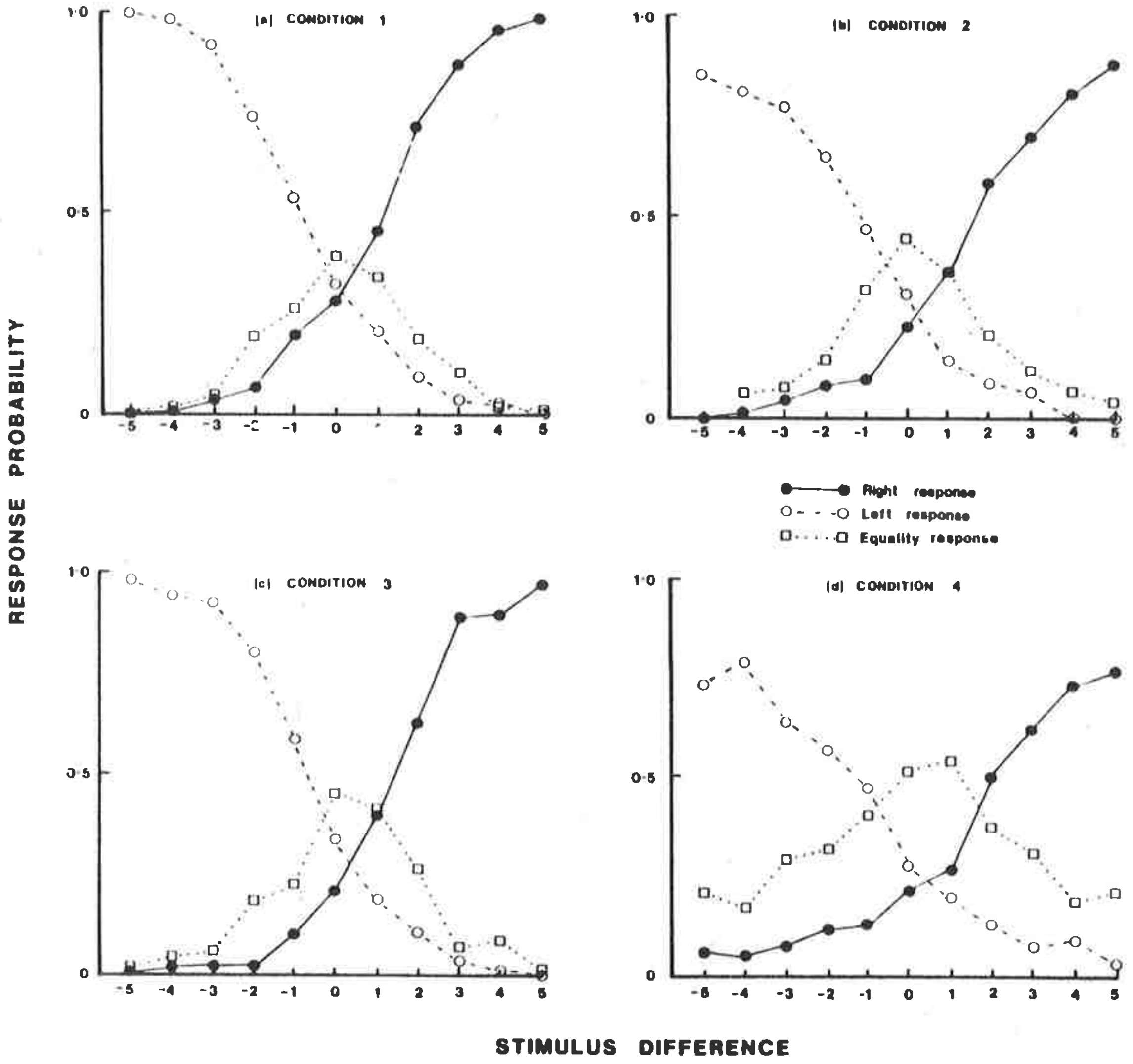


FIGURE 11. The probability of correct and incorrect responses to right stimuli (solid line), left stimuli (broken line) and equal stimuli (dotted line) at each value of stimulus difference for (a) Condition 1, (b) Condition 2, (c) Condition 3 and (d) Condition 4. Positive, non-zero values of stimulus difference indicate that the mean segment length in the right array is longer. Zero values of stimulus difference indicate that the mean segment length in the two arrays is equal. Negative, non-zero values of stimulus difference indicate that the mean segment length in the left array is longer. Probabilities were taken over the nine observers in each condition.

For example, from figure 11 it appears that the MP of equality responses in Condition 4 was higher than in Conditions 1, 2 and 3. However, a one-way analysis of variance of MP for each condition was not significant ($F(3,32) = 1.02$, $MSerror = 0.04$, $p > .05$). On the other hand, when the total probability of equality responses was considered, the analysis of variance was significant ($F(3,32) = 10.04$, $MSerror = 0.02$, $p < .001$). The probability of an equality response at any stimulus difference increased with the a priori probability of equal stimuli. The total probabilities of equality responses in Conditions 1 to 4 were 0.15, 0.30, 0.31 and 0.47, respectively.

One-way analysis of variance on the size of the IU for the four conditions produced a significant effect ($F(3,32) = 2.93$, $MSerror = 1.44$, $p < .05$). The size of the IU increased from Condition 1 to Condition 3 and decreased again in Condition 4. The mean size of the IU in Conditions 1 to 4 was 0.62, 0.78, 2.09 and 0.80, respectively. Observers in each condition were then ordered according to their MP and their IU. However, as seen in table 3, the rank-order correlation between these two orderings was not significant for any condition.

CONDITION	SPEARMAN RANK-ORDER CORRELATION COEFFICIENTS	p (ONE-TAILED)
1	0.32	> .05
2	0.18	> .05
3	0.27	> .05
4	0.02	> .05

TABLE 3. Spearman rank-order correlation coefficients and significance values relating the MP and the IU for individual observers in each condition.

Similarly, observers were ordered according to their total probability of equality responses and this ordering was correlated with the ordering according to IU. From table 4 it is apparent that, once more, the correlations were not significant and that, in this case, the correlation coefficient for Condition 4 was negative.

CONDITION	SPEARMAN RANK-ORDER CORRELATION COEFFICIENTS	P (ONE-TAILED)
1	0.28	> .05
2	0.22	> .05
3	0.25	> .05
4	-0.07	> .05

TABLE 4. Spearman rank-order correlation coefficients and significance values relating the total probability of equality responses and the IU for individual observers in each condition.

An examination of figure 11 also appears to indicate a tendency for the ogival functions for extreme responses to become flatter as the probability of equal stimuli increases. Therefore, one-way analyses of variance were performed on the measure of precision, h, of normal ogives fitted to the data for right responses and for left responses for each observer in each condition. The precision of ogives for right responses varied significantly between conditions ($F(3,32) = 3.02$, $MSerror = 0.02$, $p < .05$) while the precision of ogives for left responses did not ($F(3,32) = 2.35$, $MSerror = 0.02$, $p > .05$). However, the variation in h for right responses between conditions did not appear to be systematically related to *a priori* probability. The mean values of h for right responses for Conditions 1 to 4 were 0.40, 0.38, 0.48 and 0.28, respectively.

(b) Response time

The analysis of response time data showed a significant main effect of Discriminability (F(5,160) = 38.58, MSerror = 5553, $p < .001$) and a significant Condition x Discriminability interaction (F(15,160) = 5.53, MSerror = 5553, $p < .001$).

The Discriminability main effect arose because response time decreased as stimulus difference increased, but at a decreasing rate. The mean response times at Discriminability levels 0 to 5 were 857.95 msec, 762.85 msec, 726.75 msec, 694.59 msec, 667.93 msec and 644.21 msec, respectively.

From table 5 it can be seen that the relationship between discriminability and response time was the same in Conditions 1, 2 and 3, where equality responses were made more slowly than responses at the smallest value of stimulus difference. However, in Condition 4, times for equality responses were shorter than times for responses at the two smallest values of stimulus difference. This change in the relative times for equality and extreme responses, which gave rise to the Condition x Discriminability interaction, can be seen in figure 12 where Conditions 1 and 4 are contrasted.

CONDITION	STIMULUS DIFFERENCE					
	0	1	2	3	4	5
1	1028.71	811.90	739.23	690.47	637.41	603.16
2	793.65	721.84	676.15	657.47	643.13	630.14
3	833.03	717.37	696.92	676.50	646.86	614.79
4	776.39	800.27	794.68	753.91	744.33	728.76

TABLE 5. Mean times for correct responses at each of the six values of stimulus difference in each condition. The means were taken over the nine observers in each condition.

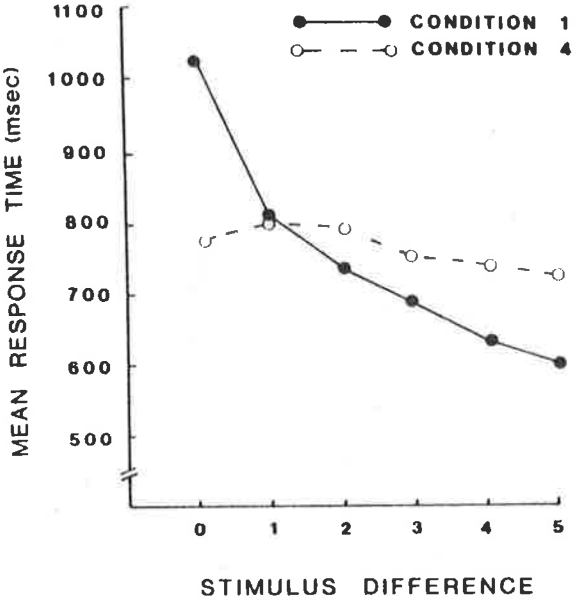


FIGURE 12. Mean times for correct responses at each of the six values of stimulus difference in Condition 1 (solid line) and Condition 4 (broken line). The means were taken over the nine observers in each condition.

Figure 13 shows the curves relating times for both correct and incorrect responses to right, equal and left stimuli and stimulus difference in each condition. In Condition 4 there were very few correct responses when the stimulus difference was small, so it was not possible to plot all points on the curves for this condition. Therefore, the pattern shown by these data is unclear and will not be discussed in detail. However, in Conditions 1, 2 and 3 there is a clear inverse relationship between times for extreme responses and stimulus difference. In all cases, the curves for equality responses lie above the curves for extreme responses, and changes in their shape across conditions appear to be unsystematic.

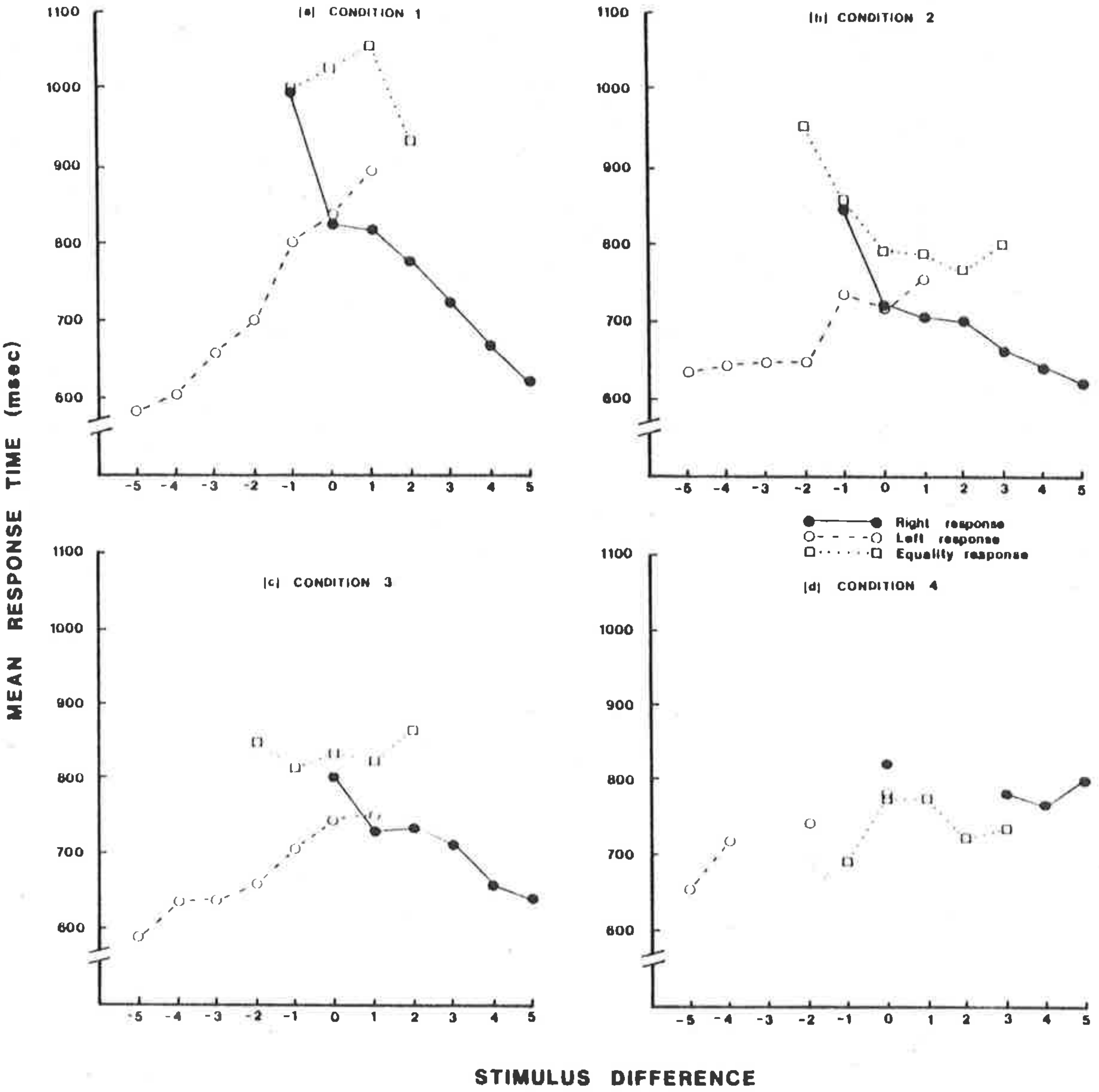


FIGURE 13. Mean times for correct and incorrect right responses (solid line), left responses (broken line) and equality responses (dotted line) at each value of stimulus difference for (a) Condition 1, (b) Condition 2, (c) Condition 3 and (d) Condition 4. Positive, non-zero values of stimulus difference indicate that the mean segment length in the right array is longer. Zero values of stimulus difference indicate that the mean segment length in the two arrays is equal. Negative, non-zero values of stimulus difference indicate that the mean segment length in the left array is longer. Times were averaged over the nine observers in each condition. Missing values arose when not all observers made a response at a given value of stimulus difference.

A two-way analysis of variance with Condition and the Correctness of responses as factors was performed on mean times for correct and incorrect responses to equal stimuli. There were repeated measures on Correctness. Both Correctness and Condition x Correctness were significant ($F(1,32) = 18.72$, $MSerror = 5488$, $p < .01$; $F(3,32) = 7.06$, $MSerror = 5488$, $p < .01$, respectively). On average, correct responses were slower than incorrect responses. The mean times for correct and incorrect responses were 857.94 msec and 782.83 msec, respectively. However, from table 6 it is apparent that the Condition x Correctness interaction was significant because this pattern only held in Conditions 1, 2 and 3, while the size of the difference between the times for correct and incorrect responses decreased from Condition 1 to Condition 3. In Condition 4 correct responses were made more quickly than incorrect responses.

CONDITION	CORRECT RESPONSES	INCORRECT RESPONSES
1	1028.79	830.62
2	793.60	719.25
3	833.03	778.39
4	776.31	803.07

TABLE 6. Mean times for correct responses and incorrect responses to equal stimuli in each condition. The means were taken over the nine observers in each condition.

The same form of analysis was carried out on mean times for correct and incorrect equality responses. The main effect of Condition was significant ($F(3,32) = 4.93$, $MSerror = 61060$, $p < .01$), but there was no consistent relationship between the mean overall times for equality responses and Condition. The mean overall times for equality responses in Conditions 1 to 4 were 1045.38 msec, 792.61 msec, 841.09 msec and 755.94 msec, respectively.

(c) Confidence

The analysis of confidence data showed a significant main effect of Discriminability (F(5,160) = 75.99, MSerror = 0.53, $p < .001$) and a significant Condition x Discriminability interaction (F(15,160) = 1.89, MSerror = 0.53, $p < .05$).

The Discriminability main effect arose because confidence increased with stimulus difference, but at a decreasing rate. The mean values of confidence at Discriminability levels 0 to 5 were 5.39, 5.99, 6.43, 7.20, 7.73 and 8.02, respectively.

From table 7 it is apparent that, in all conditions, confidence increased with stimulus difference. Trend analyses performed on each condition separately showed that, in each case, there was a significant linear trend ($p < .05$). However, the Condition x Discriminability interaction reached significance because the rate of increase in confidence was slower in Condition 4 than in any other condition. At the same time, in Conditions 1 and 2, where equal stimuli were less probable than extreme stimuli, confidence at each level of discriminability tended to be higher than in Conditions 3 and 4 where equal stimuli predominated. This is illustrated in figure 14 where Conditions 1 and 4 are contrasted.

CONDITION	STIMULUS DIFFERENCE					
	0	1	2	3	4	5
1	5.86	6.87	7.43	7.87	8.38	8.52
2	5.86	6.40	7.01	7.92	8.36	8.65
3	4.23	4.80	5.57	6.48	7.39	7.69
4	5.61	5.88	5.70	6.51	6.78	7.22

TABLE 7. Mean confidence in correct responses at each of the six values of stimulus difference in each condition. The means were taken over the nine observers in each condition.

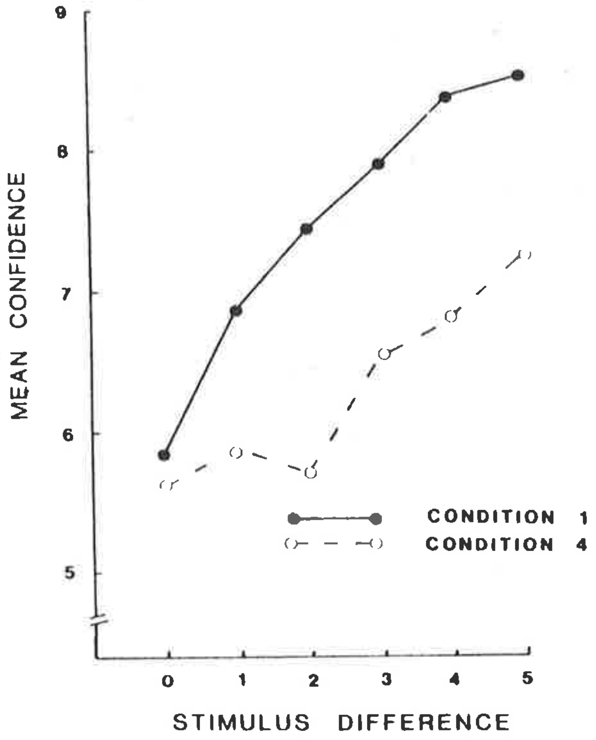


FIGURE 14. Mean confidence in correct responses at each of the six values of stimulus difference in Condition 1 (solid line) and Condition 4 (broken line). Confidence ratings were averaged over the nine observers in each condition.

Figure 15 shows the curves relating confidence for both correct and incorrect responses to right, equal and left stimuli and stimulus difference in each condition. In all conditions there was a direct relationship between confidence in extreme responses and stimulus difference. This trend was evident even in Condition 4 where the data for several points were not sufficient to allow them to be plotted. In general, in all conditions the curves for equality responses were fairly flat.

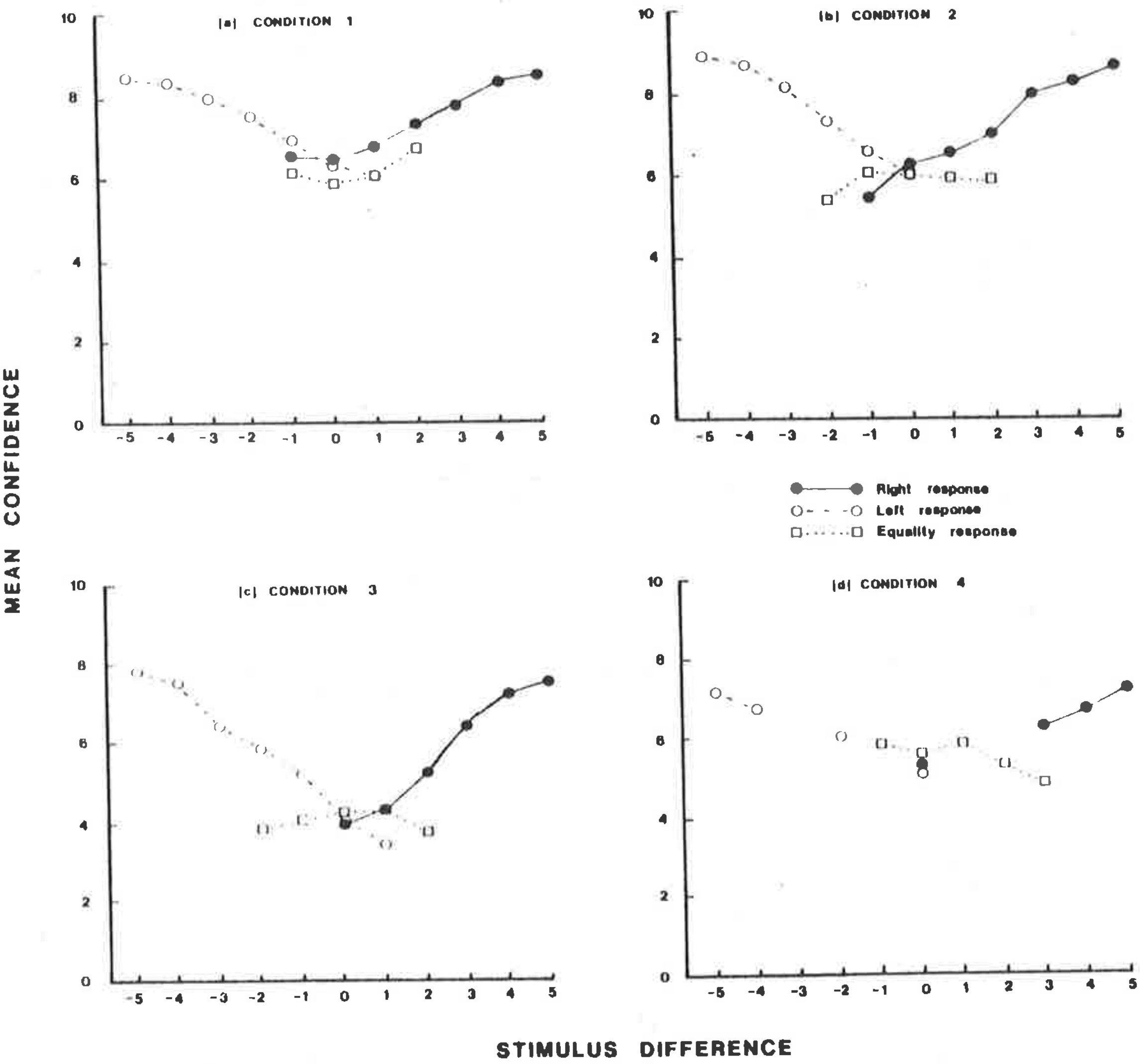


FIGURE 15. Mean confidence in correct and incorrect right responses (solid line), left responses (broken line) and equality responses (dotted line) at each value of stimulus difference for (a) Condition 1, (b) Condition 2, (c) Condition 3 and (d) Condition 4. Positive, non-zero values of stimulus difference indicate that the mean segment length in the right array is longer. Zero values of stimulus difference indicate that the mean segment length in the two arrays is equal. Negative, non-zero values of stimulus difference indicate that the mean segment length in the left array is longer. The confidence ratings were averaged over the nine observers in each condition. Missing values arose when not all observers made a response at a given value of stimulus difference.

A two-way analysis of variance with Condition and Correctness of responses as factors (with repeated measures on Correctness) was performed on mean confidence in correct and incorrect responses to equal stimuli. Both Condition and Correctness main effects were significant ($F(3,32) = 2.93$, $MSerror = 3.78$, $p < .05$; $F(1,32) = 5.33$, $MSerror = 0.18$, $p < .05$, respectively) but their interaction was not significant. There was no simple relationship between overall confidence and Condition, with the mean overall confidence in Conditions 1 to 4 taking the values 6.14, 5.92, 4.12 and 5.42, respectively. Confidence in responses to equal stimuli decreased with an increase in the proportion of equal stimuli from Condition 1 to Condition 3 and then increased when equal stimuli became more probable than extreme stimuli in Condition 4. The Correctness effect arose because confidence in correct responses was higher than confidence in incorrect responses. The mean confidence for correct and incorrect responses to equal stimuli was 5.39 and 5.16, respectively.

The same form of analysis was then carried out on confidence in correct and incorrect equality responses. Only Condition was significant ($F(3,32) = 4.28$, $MSerror = 3.45$, $p < .05$) and mean overall confidence values followed the same pattern across conditions as described above. The mean overall confidence in Conditions 1 to 4 was 5.81, 5.67, 4.11 and 5.52, respectively.

4. Discussion

(a) Accuracy

The finding that the accuracy of responses to equal stimuli increased and the accuracy of responses to extreme stimuli decreased as the probability of equal stimuli increased, agrees with Vickers' (1975) findings and can be accounted for by the model based on Signal Detection Theory, the random walk and clock model or the accumulator model. In the model based on Signal Detection Theory, an increase in the a priori probability of equal stimuli will give rise to a change in the position of the cutoff, as determined by the ideal observer hypothesis, in such a way that more equality responses will be given overall. Therefore, the proportion of equal stimuli which are identified correctly will increase. In the random walk and clock model, t will decrease as the a priori probability of equal stimuli increases. Once more, this will result in an increase in the total proportion of equality responses, and therefore the number made accurately should increase. Similarly, in the accumulator model, when x_e is relatively low as a result of bias towards equality responses induced by an increase in the probability of equal stimuli, equality responses will occur more often overall, and therefore the accuracy rate should increase.

The observed psychometric functions, averaged over all observers, are of the same general form as those found by Kellogg (1931) and Vickers (1975) and are predicted by each of the models. However, an examination of the properties of the individual psychometric functions of each observer in each condition provided little support for previous findings. For example, while Urban (1910) showed an increase in IU with an increase in the a priori probability of equal stimuli, there was no consistent variation in IU in the present data. In particular, IU increased from Condition 1 to Condition 3 but then decreased in Condition 4 where equal stimuli were most probable. In addition, Urban (1910) and Vickers (1975) found that observers could be ordered according to their MP or their IU, and that there was a perfect positive correlation between the rank orderings of observers on these two measures. There was no evidence of this correlation in the present data.

Similarly, from Urban's (1910) results it would be expected that the MP of equality responses would increase with the a priori probability of equal stimuli. Once more this was not the case in these data. In contrast, there was a consistent increase in the total proportion of equality responses from Condition 1 to Condition 4. As seen above, this would be expected on the basis of any of the models. However, when observers were ranked according to this measure of the probability of equality responses and these rankings were correlated with rankings of IU, there was no significant relationship.

The increase in the total proportion of equality responses as the probability of equal stimuli increased could be interpreted as evidence for a tendency for a systematic change in the precision of the ogives for extreme responses. Precision (h) should decrease with an increase in the *a priori* probability of equal stimuli. This was not supported by the statistical analysis which showed a significant, but unsystematic, variation in h for right responses and no change in h for left responses. Therefore, it appears that the increase in the total proportion of equality responses represented instead a movement of the points along the same ogive.

Although the overall analysis of accuracy data appeared to indicate that bias towards equality responses increased systematically from Condition 1 to Condition 4, this was not supported by the examination of the measures of MP, IU and h. For these measures, bias towards equality responses appeared to increase from Condition 1 to Condition 3, but in Condition 4 there was evidence of a bias against equality responses. IU decreased and h for right responses was very low in Condition 4. Since observers were not told the actual stimulus probabilities, it is possible that, in Condition 4, their subjective estimates of the probabilities were not as extreme as the actual probabilities. As a result, they may have considered that they were making more equality responses than were appropriate and therefore, they may have tried to reduce this apparent discrepancy by adopting a bias against these responses.

(b) Response time

The finding that times for correct equality responses decreased to be shorter than times for correct extreme responses as the a priori probability of equal stimuli increased provides support for the findings of Vickers (1975). This finding can be accommodated by the model based on Signal Detection Theory by means of the choice of an appropriate latency function. However, the random walk and clock model cannot account for the finding that times for equality responses were both shorter and longer than those for extreme responses. In contrast, simulations of the accumulator model show a decrease in the times for equality responses below the times for some extreme responses when bias towards equality responses is increased (Vickers, 1979). On the other hand, these simulations also show a consistent decrease in the times for equality responses with an increase in the bias towards these responses. This trend was not apparent in the present data.

The inverse relationship which was observed between response time and stimulus difference for equality and extreme responses in Conditions 1, 2 and 3 can be accounted for by each of the models. These results can be accommodated by the model based on Signal Detection Theory if an appropriate latency function is adopted. Alternatively, in the random walk and clock model, the assumption that t is higher than the times for extreme responses would give rise to this pattern of data. Similarly, these findings may be explained by an accumulator model where observations of a more discriminable stimulus will provide more evidence towards a response. Therefore, fewer observations will be needed to satisfy the criterion, so times will be shorter.

An examination of the curves relating response time and stimulus difference showed that the form of the curves was the same as those found by Kellogg (1931) and Vickers (1975). As in Vickers' (1975) study, there was evidence that the position of the curve for equality responses changed with a priori probability. As the a priori probability of equal stimuli increased, the curve appeared to move downwards. This is consistent with all of the models. However, in contrast to Vickers' (1975) findings, the shape of the curve for equality responses did not appear to vary systematically with the a priori probability of equal stimuli. Instead, in all cases, times for correct equality responses tended to be slower than times for incorrect equality responses.

(c) Confidence

The finding that, in each condition, the direct relationship between confidence and stimulus difference for equality and extreme responses showed a significant linear trend, implies that confidence in equality responses lies on a continuum with confidence in extreme responses. Therefore, these data do not provide any support for the suggestion that equality judgments fall on a qualitatively different confidence scale (e.g. Fernberger, 1930; Kellogg, 1931; Fernberger & Irwin, 1932).

Neither the model based on Signal Detection Theory nor the random walk and clock model can account for these changes in confidence with the changes in stimulus difference. In the Signal Detection Theory model, the inclusion of Audley's (1960) hypothesis to account for confidence would result in the prediction that confidence would vary inversely with response time. This inverse relationship did not hold in Condition 4, where times for equality responses were short but confidence was low. Alternatively, if Audley's (1960) hypothesis is incorporated into the random walk and clock model in which t is constant, confidence in both correct and incorrect equality responses should always be lower than confidence in extreme responses. In these data, this was true only for Condition 1. When it is assumed that t varies randomly, either confidence in equality responses should always be lower than confidence in extreme responses or it should always be higher. This prediction also failed to find any support in Conditions 2, 3 and 4 in the present data where confidence in equality responses lay between the values of confidence for correct and incorrect extreme responses.

On the other hand, the accumulator model predicts an inverse relationship between confidence and stimulus difference for extreme responses. However, according to the accumulator model, confidence in equality responses should be directly related to the probability of equal stimuli. This was not apparent in these data. Simulations of the accumulator model show sigmoidal confidence functions for extreme responses as seen in the present data, but, according to this model, the curves for equality responses tend to be flat with a bow upwards (Vickers, 1975). Although the observed curves were relatively flat, there was no consistent bowing between conditions. This was supported by the statistical comparison of confidence in correct and incorrect equality responses.

5. Summary and Conclusions

The major findings from this experiment can be accommodated by all of the models provided that it is assumed that there are systematic changes in criterion values. Nevertheless, some of the indices which might have been assumed to provide evidence of criterion change showed less systematic variation.

The effects of the manipulation of a priori probability were evident to some extent in all three dependent variables. For example, while there was no difference in the accuracy measures in Conditions 1, 2 and 3, in Condition 4, accuracy for extreme responses was lower than in the other conditions and accuracy for equality responses was higher than in the other conditions and higher than accuracy for responses at the smallest stimulus difference in this condition. Therefore, it appears that an increase in the a priori probability of equal stimuli tended to result in an increase in the accuracy of equality responses and a decrease in the accuracy of extreme responses. The psychometric functions plotted from these data were of the same form as previously observed empirical curves. In addition, there was an increase in the total proportion of equality responses as the proportion of equal stimuli increased.

At the same time, in each condition there was an inverse relationship between response time and discriminability. In particular, the longest times were for equality responses in all except Condition 4. It appears that when the bias towards equality responses was high, times for these responses tended to decrease. This finding received further support from the examination of curves relating times for correct and incorrect responses and stimulus difference. The curve for equality responses appeared to move downward as the a priori probability of equal stimuli increased.

In addition, for all conditions, confidence decreased with stimulus difference and was lowest for equality responses. However, once more Condition 4 appeared to show evidence of the effects of bias towards equality responses. In this condition, the difference between confidence in equality responses and confidence in responses at the smallest stimulus difference was markedly smaller than for the other conditions.

Although there were these indications of the effects of bias, several measures did not show any systematic changes. For example, there was no consistent change in the MP or the IU of the psychometric functions from Condition 1 to Condition 4. Similarly, the precision of the ogives for extreme responses did not vary systematically. At the same time, there was no evidence of any variation in the shape of curves relating times for equality responses and stimulus difference. In the same way, an examination of the curves relating confidence in correct and error responses and stimulus difference for each condition showed that there was no regular variation in the shape or the relative position of the curves when a priori probability was changed.

However, overall, it is Condition 4 which showed inconsistent effects of bias. In most cases there were regular changes in the response measures from Condition 1 to Condition 3, but Condition 4 did not follow the same pattern. It appears that the use of a very high probability for equal stimuli in Condition 4 may have resulted in observers adjusting their performance to match much less extreme subjective estimates of stimulus probability.

On the basis of these data alone, some measures do not appear to be very good indicators of bias. However, possible reasons for this appear to arise from limitations in several aspects of the experiment. For example, since this study was performed to examine the effects of a priori probability, observers were not given any information about the stimulus probability in instructions. As a result, the contrast between the findings of Vickers (1975) and the present experiment may be interpreted as indicating that the manipulation of a priori probability, without informing observers of the actual stimulus probability, does not have a strong biassing influence on responding. Alternatively, as suggested above, this may have resulted in observers forming inaccurate assessments of the stimulus probabilities which may have influenced their performance.

In addition, it appears that observers found this form of discrimination task very difficult. This is suggested by the large number of practice trials required to reach stable performance and the low accuracy rates overall. The inclusion of several levels of discriminability may have increased the complexity of the demands placed on the observers. Furthermore, the use of separate groups of observers for the different bias conditions may have resulted in individual differences masking any variation between the groups.

B. EXPERIMENT 2

1. Introduction

In Experiment 2, the effects of bias on accuracy, time and confidence in a three-category task were examined using a simpler task (Vickers, Nettelbeck & Willson, 1972; Vickers *et al.*, 1977), and a different design. As in the previous experiment, bias was varied by means of changes in the *a priori* probabilities of the three stimuli. However, since Condition 4 of Experiment 1, in which equal stimuli were much more probable than extreme stimuli, appeared to produce anomalous results, the two values of *a priori* probability used in this experiment were not as extreme as in that condition. In addition, in this case, each observer performed under the two different probability conditions within a single experimental session. The use of a within-subjects design meant that the effects of individual differences were not confounded with the effects of bias. To simplify the task further, in this experiment discriminability was held constant. At the same time, the confidence rating scale was changed to a five point scale and instructions for its use were altered in an attempt to improve the consistency of observers' ratings.

2. Method

(a) Stimuli

The stimuli consisted of two vertical lines 40 mm apart, with their upper ends terminated by a horizontal line (see figure 16). These pi-shaped figures were drawn in white on the medium grey background of the video monitor. All lines were approximately 0.5 mm wide. The horizontal line was of constant length (80 mm) and always appeared in the same position on the screen. The vertical lines were either equal in length or of different lengths. When equal in length, both vertical lines measured 187 mm. When of different lengths, the longer line, which appeared on either the right or the left side, measured 187 mm and the shorter line measured 185.5 mm. This difference in length subtended a visual angle of less than 0.1°, when viewed from a distance of four metres.

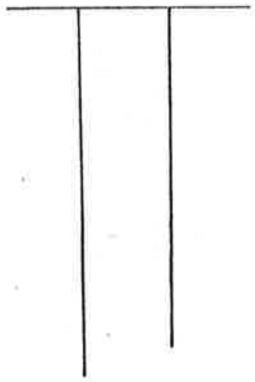


FIGURE 16. A drawing of the type of stimulus used in Experiment 2. The horizontal line was 80 mm long. When the vertical lines were of different lengths, as represented here, the longer vertical line measured 187 mm and the shorter vertical line measured 185.5 mm. When equal in length, both vertical lines measured 187 mm. In the drawing the left vertical line is longer.

(b) Apparatus

A PDP 8/F digital computer under the control of a second disk-based PDP 8/E digital computer generated the stimuli on a video monitor. It determined whether the vertical lines would be of equal or different lengths, and the side on which the longer line would occur when the lines were different in length.

The observer was seated at a table on which were eight identical microswitched perspex buttons, each 2 mm in diameter, and a perspex marker. The buttons were in two groups. To indicate the side on which the longer line occurred, or that the lines were equal in length, the observer pressed one of three buttons which were centrally located on the table. The three buttons were in a semicircle, adjacent buttons being 3 cm apart, centre to centre, and all were equidistant from the perspex marker which was about 5 cm closer to the observer. Confidence was rated by pressing one of five buttons which formed a line close to, and parallel with the further edge of the table. These buttons were each 3 cm apart, centre to centre. There was an easily distinguishable white spot on the table immediately below the centre button of the row of five. Depression of any button by about 1.5 mm was necessary to make contact which was accompanied by a discernible 'click'.

The PDP 8/E computer measured the time between stimulus onset and the response, and recorded and classified responses and confidence ratings.

(c) Observers

The 20 observers, 10 men and 10 women, were students enrolled in the first year psychology course at the University of Adelaide whose participation was credited towards a course requirement. They ranged in age from

17 to 26 years and all declared themselves to be right-handed. All observers were naive with respect to the aims of the experiment.

(d) Design

Two groups of ten observers were formed, each comprising five men and five women. The groups performed under different order conditions, observers being tested individually for approximately one hour.

In both order conditions, stimuli were presented in 10 blocks: 4 practice blocks each of 60 trials preceded 6 experimental blocks each of 120 trials. For both groups of observers, the occurrence of a stimulus in which the right line was longer, the left line was longer, or the lines were equal in length, was equiprobable (p = 0.33) in each practice block and the first experimental block. The remaining five experimental blocks took one of two forms:

A. The probability of an equal stimulus was 0.16 and right and left stimuli were equiprobable with a probability of 0.42.

B. The probability of an equal stimulus was 0.68 and right and left stimuli were equiprobable with a probability of 0.16.

One group of observers, Order AB, performed the five blocks in order BABAB, while the other group of observers, Order BA, performed the five blocks in order ABABA. In both cases, the first of these experimental blocks was treated as practice and only the last four blocks were considered in the statistical analyses.

Practice blocks were separated by short rest periods in which observers were given the opportunity to clarify instructions. Experimental trials occurred in an uninterrupted sequence in which there was no indication of the changes in stimulus probability which were taking place.

(e) Procedure

The experiment was performed in a dimly lit room. Approximately five minutes were allowed for adaptation to these conditions before commencing. The observer, seated at the table of response buttons, was four metres directly in front of the video screen.

After the form of the stimulus was explained to the observer, he was instructed that, for each stimulus presented, he should indicate the side on which the longer line occurred or that the lines were equal in length by pressing the appropriate response button, the left button if the left line were judged longer, the centre button if the lines were judged equal in length or the right button if the right line were judged longer. The observer was instructed that, following each such response, he must rate his confidence in its accuracy by pressing one of the row of five buttons. Before the practice trials began it was explained that the left-most button represented the lowest confidence rating, 1, and the right-most button the highest rating, 5, with the intervening buttons completing a regular linear scale between these extremes. At the completion of the second block of practice trials, the observer was instructed that, when rating his confidence on all future trials, he should use the centre confidence button, button 3 (labelled with a white spot), to indicate his average level of confidence as experienced to this point. As before, lower ratings would be represented by those buttons to the left on the table, buttons 2 and 1, while higher ratings would be given on buttons to the right, buttons 4 and 5.

The observer was required to use the index finger of his right hand to respond. On each trial the observer should move his finger from the marker to press first a response button and then a confidence button, returning his finger to the marker before the start of the next trial.

3. Results

Analyses of variance with Order (AB or BA, as defined in section B.2(d)), Period (first two experimental blocks or second two experimental blocks), Probability (block type A or type B, as defined in section B.2(d)), Half (first or second half of an experimental block) and Response (right, equality or left) as factors were used to analyse data from the last four experimental blocks. Analyses of variance were carried out according to a crossed factorial design with repeated measures on Period, Probability, Half and Response. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses.

(a) Accuracy

The analysis of accuracy data showed significant main effects of Period ($F(1,18) = 10.14$, $MSerror = 0.14$, $p < .001$) and Response ($F(2,36) = 12.39$, $MSerror = 0.72$, $p < .001$) and a significant Order x Probability x Response interaction ($F(2,36) = 5.16$, $MSerror = 0.15$, $p < .05$).

The Period main effect reached significance because, for both groups of observers, accuracy was higher in Period 1 than in Period 2. The percentages of stimuli for which responses were correct in Periods 1 and 2 were 91.52% and 88.86%, respectively.

The Response main effect arose because both right responses and left responses were significantly more accurate than equality responses, while there was no difference between the accuracy of right and left responses. The percentages of right, equal and left stimuli for which responses were correct were 93.68%, 80.61% and 92.40%, respectively.

The interaction between Order, Probability and Response was significant because, in all cases except blocks of type B in Order BA, accuracy for both extreme responses was significantly higher than accuracy for equality responses, while there was no difference between the accuracy for responses to right and left stimuli. However, in blocks of type B in Order BA there was no significant difference between the accuracy of the three responses. From figure 17 it is apparent that the accuracy of equality responses in this condition was significantly higher than in blocks of type B in Order AB and blocks of type A in either order condition.

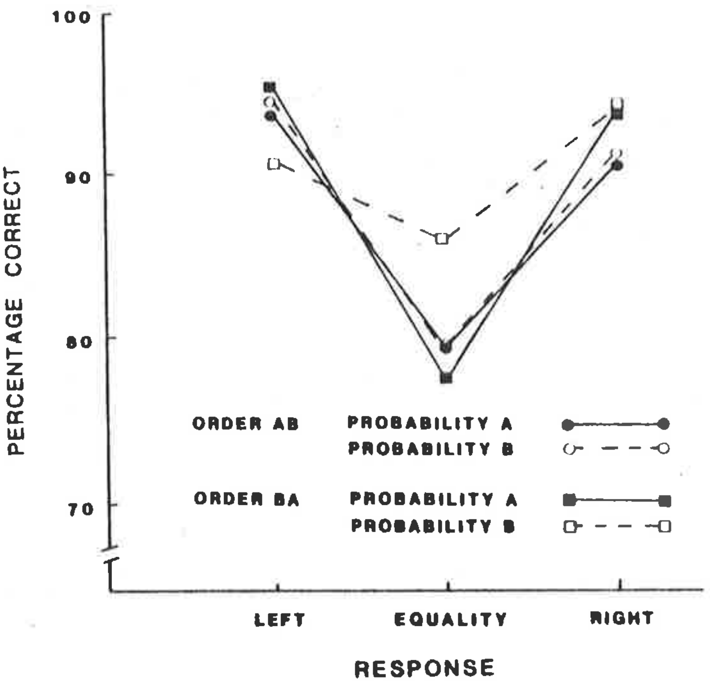


FIGURE 17. The percentage of left, equal and right stimuli for which responses were correct for Probability A (solid line) and Probability B (broken line) in Order AB (circles) and in Order BA (squares). The percentages were taken over the 10 observers in each order.

(b) Response time

The analysis of response time data showed a significant main effect of Response (F(2,36) = 6.63, MSerror = 87188, $p < .01$) and significant interactions of Probability x Response (F(2,36) = 9.91, MSerror = 18422, $p < .001$) and Period x Half x Response (F(2,36) = 4.11, MSerror = 16393, $p < .05$).

The Response main effect arose because mean times for equality responses were significantly longer than mean times for both right and left responses, while there was no difference between times for right and left responses. The mean times for right, equality and left responses were 915.30 msec, 1033.09 msec and 958.23 msec, respectively.

The pattern of times for each of the three responses for Probability A and Probability B is shown in figure 18. The significant Probability x Response interaction arose because, although equality responses were slower than both

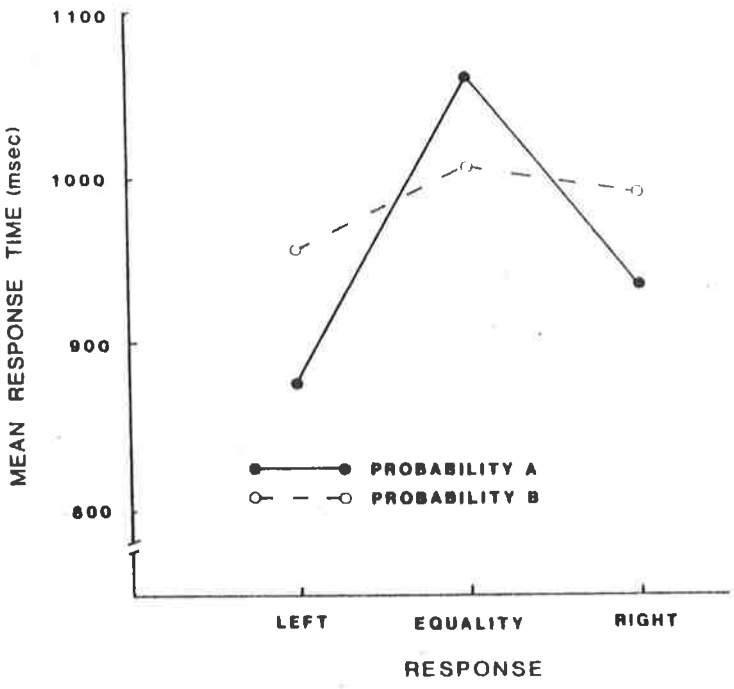


FIGURE 18. Mean times for correct left, equality and right responses for Probability A (solid line) and for Probability B (broken line). The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

right and left responses in both probability conditions, this difference reached significance only in Probability A. In addition, from the figure it is apparent that the mean times for each of the three responses were shorter when the corresponding stimulus had a high probability of occurrence than when the stimulus was less likely to occur. In particular, mean times for equality responses were shorter in Probability B than in Probability A, while mean times for right and left responses were shorter in Probability A than in Probability B.

The Period x Half x Response interaction was significant because left responses were significantly longer in the first half of blocks in Period 1 than at any other time (see figure 19). As a result, in the first half of

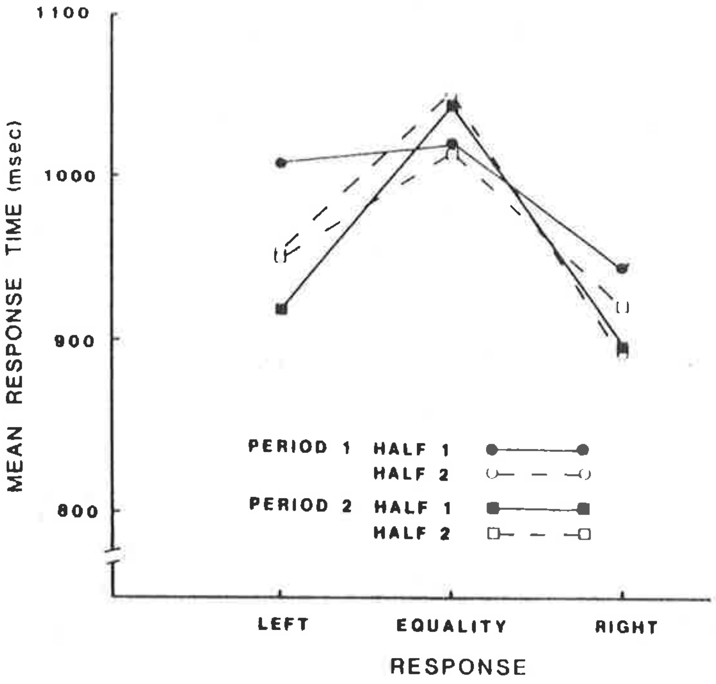


FIGURE 19. Mean times for correct left, equality and right responses in Half 1 (solid line) and Half 2 (broken line) of Period 1 (circles) and of Period 2 (squares). The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

blocks in Period 1 there was no difference between the times for equality and left responses which were both significantly longer than times for right responses. In all other cases, equality responses were significantly slower than both extreme responses and there was no difference between the times for the two extreme responses.

(c) Confidence

The analysis of confidence data showed a significant main effect of Response (F(2,36) = 4.70, MSerror = 1.05, $p < .05$) and a significant Probability x Half x Response interaction (F(2,36) = 3.90, MSerror = 0.11, $p < .05$).

The Response main effect arose because, while there was no difference between confidence in right and left responses, confidence in both of these extreme responses was significantly higher than confidence in equality responses. The mean confidence in right, equality and left responses was 3.58, 3.29 and 3.58, respectively.

The Probability x Half x Response interaction is shown in figure 20. In blocks of type A, confidence in equality responses increased from the first to the second half of blocks while confidence in extreme responses decreased from the first to the second half of blocks. In contrast, in blocks of type B, confidence in equality responses decreased from the first to the second half of blocks while confidence in extreme responses increased from the first to the second half of blocks. In addition, there were changes in confidence between blocks. Confidence in equality responses increased while there was no change in confidence in right and left responses from Half 2 of blocks of type A to Half 1 of blocks of type B. On the other hand, confidence in equality responses decreased while confidence in right and left responses increased from Half 2 of blocks of type B to Half 1 of blocks of type A. However, none of these differences was significant.

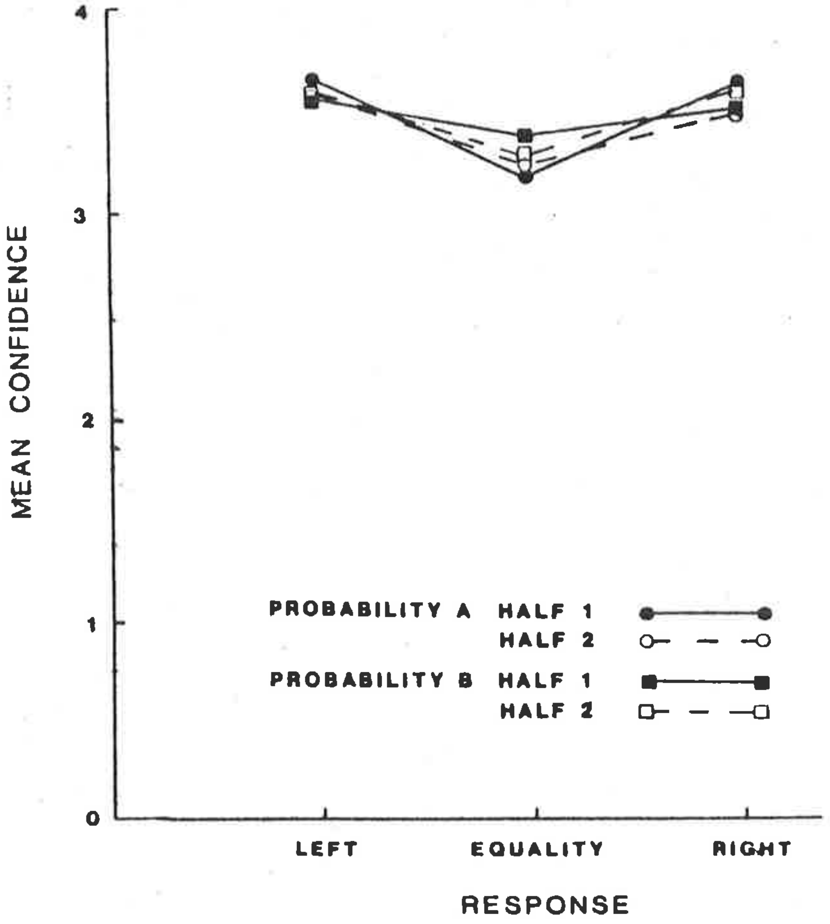


FIGURE 20. Mean confidence in correct left, equality and right responses in Half 1 (solid line) and Half 2 (broken line) of blocks of Probability A (circles) and blocks of Probability B (squares). The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

(d) Confidence and response time

To this point the analysis has examined only those relationships exhibited between probability conditions. Consideration of the data within probability levels provided information on the form of the relationship between confidence and response time.

For each observer, for each block of trials, and for each response, confidence ratings for correct and incorrect responses were plotted against the corresponding response time values. From the total of 240 plots, 191 of the slopes of the best fitting straight lines were negative ($p < .001$, binomial test), showing an inverse relationship between confidence and response time.

PROBABILITY	RESPONSE	NO. OF NEGATIVE SLOPES	p
A	RIGHT	31	< .001
	EQUALITY	27	< .05
	LEFT	35	< .001
B	RIGHT	29	< .005
	EQUALITY	35	< .001
	LEFT	32	< .001

TABLE 8. The number of negative slopes obtained when, for each observer, confidence ratings were plotted against response time values and straight lines were fitted. Separate lines were fitted to data from each response in each block of trials for each of the 20 observers so that the maximum possible number of negative slopes for each response in each probability condition was 40. The associated binomial probability values are also shown.

This inverse relationship was still evident when each response was considered separately. In this case, from the total of 80 plots for each response, the numbers of negative slopes for right, equality and left responses were 60, 62 and 67, respectively (in all cases, $p < .001$, binomial test). The predominantly inverse relationship was also evident when each response was considered in each probability condition (see table 8).

4. Discussion

(a) Accuracy

The finding that responses to equal stimuli were less accurate than responses to extreme stimuli in both conditions appears to imply that observers were biassed against equality responses throughout the experiment, irrespective of the a priori probability of equal stimuli. This tendency was noted previously by Hayden (1906) and Angell (1907). Each of the models can accommodate this finding. For example, in the model based on Signal Detection Theory it would be assumed that bias against equality responses would produce a change in the value of the cutoff as determined by the ideal observer hypothesis as a result of changes in the subjective costs and values of the three responses. The cutoff would move to a point where few equality responses would be made and therefore, the overall accuracy of equality responses would tend to be low. Since costs and payoffs and a priori probabilities may vary independently in the ideal observer hypothesis, this bias against equality responses is independent of any effect of the manipulation of stimulus probability. According to the random walk and clock model, t would be high, so few equality responses would be made and overall accuracy would be low. In the accumulator model, x_e would be high relative to x_g and x_l so, once more the accuracy of equality responses would tend to be low.

Since the Period main effect in accuracy data was not accompanied by corresponding effects on response time or confidence, it seems unlikely that it resulted from criterion change during the experiment. Instead it may be associated with a progressive decrease in attention or increase in fatigue. Therefore, although this result is not inconsistent with any of the models, it is not a simple prediction of any of them.

The Order x Probability x Response interaction appears to indicate that the biassing effect of _a priori_ probability was evident only in Order BA. In Order AB, the accuracy of responses to the three stimuli did not change from Probability A to Probability B. In both conditions, equality responses were less accurate than extreme responses. In contrast, in Order BA, there was a significant increase in the accuracy of equality responses from Probability A to Probability B. In Probability A, where equal stimuli had a low probability, equality responses were less accurate than both extreme responses, while in Probability B, where equal stimuli had a high probability, there was no difference in accuracy for equality and extreme responses. It appears that, on the basis of the initial practice trials, observers may have formed the expectation that the three stimuli would be equiprobable. If observers subsequently experienced a block in which equal stimuli had a high probability first, as in Order AB, they may have found that they were making many more equality responses than expected. Therefore, in order to reduce the apparent discrepancy between their expectations and their responding, they may have adopted a bias against equality responses from that point on in the experiment. In this case, equality responses should be less accurate than extreme responses overall, as seen in Order AB in the present data. On the other hand, if observers first experienced a block in which equal stimuli had a low probability, as in Order BA, they would have made fewer equality responses than expected and therefore, they may have adopted a bias towards equality responses. Such a bias would be more apparent when equal stimuli were more probable in blocks of Probability B, as seen in these data. This effect does not correspond to a simple prediction of any of the models but it is not inconsistent with any of them.

(b) Response time

The finding of longer times for equality responses than for extreme responses confirms one of the earliest findings in the study of three-category judgment (e.g. Hayden, 1906; Angell, 1907; George, 1917), and appears to imply that observers were biassed against equality responses throughout the experiment. In the Signal Detection Theory model the choice of an appropriate latency function would account for this relationship. Alternatively, the assumption of a high value for t in the random walk and clock model, or of x_e in the accumulator model would lead to the prediction of this finding.

However, the Probability x Response interaction implied that the manipulation of a priori probability was influencing the degree of bias against equality responses. When equal stimuli had a low probability, times for equality responses were significantly longer than times for left and right responses. On the other hand, when equal stimuli had a high probability, although times for equality responses were longer than times for left and right responses, this difference was not significant. In other words, there appeared to be a greater or lesser degree of bias against equality responses depending on a priori probability. This result can be explained in terms of the Signal Detection Theory model by the choice of a suitable latency function. Alternatively, in terms of the random walk and clock model, if it is assumed that t decreases as the a priori probability of equal stimuli increases, times for equality responses should decrease as observed in the present data. At the same time, the model predicts that the times for equality responses will always be longer than times for extreme responses. This pattern is suggested by the mean values shown in figure 12, but an examination of the raw data showed that, when times for individual responses were considered, this was not always the case. On the other hand, in the accumulator model it is assumed that the criterion for equality responses decreases as the result of an increase in the probability of equal stimuli, while the criteria for extreme responses increase. Therefore, times for equality responses should decrease overall and times for extreme responses should increase. This is the pattern seen in the present data.

The Period x Half x Response interaction did not appear to be indicative of the biassing effects of a priori probability and, although this finding does not necessarily contradict any of the models, it does not correspond to a simple prediction of the models as they stand.

(c) Confidence

The finding that overall, confidence in equality responses was lower than confidence in extreme responses appears to reflect the previous observations that equality responses involve doubt more often than extreme responses (e.g. Fernberger, 1930; Fernberger & Irwin, 1932). Once more this seemed to imply that observers were biassed against equality responses. This finding is consistent with all of the models. Since confidence is varying inversely with response time, Audley's (1960) hypothesis in combination with either the model based on Signal Detection Theory or the random walk and clock model would predict this result. Alternatively, in the accumulator model when the observer is biassed against equality responses, the value of x_e is assumed to be low. Therefore, the maximum possible confidence in equality responses will be low.

As there is no comparable finding in the response time data, the Probability x Half x Response interaction cannot be explained by either the Signal Detection Theory model or the random walk and clock model and Audley's (1960) hypothesis. As stated by Vickers (1979), the accumulator model predicts higher confidence for the less probable response, since the criterion for this response should be high, so that when it is made, the difference in the accumulated totals should be at a maximum. In the present data, this pattern was found in the second half of blocks. Confidence in equality responses increased from Half 1 to Half 2 of blocks of type A when equal stimuli had a low probability, and decreased from Half 1 to Half 2 of blocks of type B when equal stimuli had a high probability. Confidence in extreme responses varied in the opposite way. On the other hand, the measures in the first half of blocks did not appear to be consistent with the accumulator model. However, since the changes in confidence across block boundaries were in the same direction as the change within the preceding block, it appears that there may have been a lag in the change in performance in response to the change in probability. Indeed, in this experiment, since observers were not told about the manipulation of a priori probability it is possible that they would have become aware of the change only after several trials. In this case, these data would be consistent with an accumulator model in which there was a lag in the adjustment of the target levels of confidence.

(d) Confidence and response time

From these data it is clear that there was an inverse relationship between confidence and response time within bias conditions. This provides support for Audley's (1960) proposal that confidence is an inverse function of the number of observations needed for a decision. Therefore, this finding may be accommodated by the model based on Signal Detection Theory or the random walk and clock model in combination with Audley's (1960) hypothesis. The inverse relationship between confidence and time is also predicted by the accumulator model.

5. Summary and Conclusions

The majority of the findings from this experiment are consistent with all of the models in which they would be mediated by changes in criterion values. However, the effects of variations in a priori probability appeared to vary between response measures, and, as in Experiment 1, observers appeared to be biassed against equality responses overall, so that the effects of the experimental manipulation were less clear.

Although the findings from the present experiment were by no means conclusive, there appeared to be some evidence of the influence of bias due to changes in a priori probability in each of the response measures. The accuracy data from Order BA showed clear changes in performance as a priori probability was varied as predicted by all of the models. Responses to the more probable stimulus were more accurate relative to responses to the less probable stimulus. In contrast to the predictions of all of the models, accuracy in Order AB did not change between conditions. In response time data, times were shorter for responses to the more probable stimulus. This was consistent with all of the models. However, the pattern of variation shown by confidence measures contrasted with that shown by accuracy and response time data. The influence of changes in a priori probability on confidence continued throughout the subsequent block, with confidence changing from the first to the second half of the block. In particular, it appeared that the measure in the first half of the block represented a continuation of the change in performance from the previous block, and that the effects of the change in a priori probability did not become apparent until the second half of the block. This might be expected in the situation in which observers were not told that a priori probability would change. Since there was no corresponding finding in response time data, this effect was not consistent with Audley's (1960) hypothesis applied to either

Signal Detection Theory or the random walk and clock model. However, if the confidence measures in the second half of the block did represent the response to the change in a priori probability, these data were consistent with an accumulator model.

From the data from both Experiments 1 and 2 it appears that the manipulation of a priori probability in a three-category task does not induce a strong bias in responding. One possible reason for this may be that the observers, who were all enrolled in courses at a university, may be trained to look for differences. In this way they may be biassed towards extreme responses prior to the experiment. Furthermore, previous studies have shown that there are large individual differences in willingness to make equality responses. This variability, coupled with a trained reluctance to make equality responses may have swamped the effects of a priori probability. In addition, as mentioned above, in contrast to Vickers' (1975) study, no instructions were given about the changes in a priori probability. In the light of Vickers' (1975) results it appears that instructions may have stronger effects on criterion change than do a priori probabilities. Indeed, this would appear to support the notion that the observer's assessment of probabilities may influence responding.

In order to overcome these problems it appears necessary to change some aspects of the experimental task and procedure. For example, the observer's natural reluctance to make equality responses could be counteracted either by changing the instructions or by using a task in which equality responses are not included.

Since the effects of a priori probability alone are of interest here, a task which does not require equality responses will be used. Furthermore, it appears that it would be better to use a task in which the observer's estimates of a priori probability were more equal and where the pre-existing bias towards any one response was less marked.

CHAPTER 4

A. EXPERIMENT 3

1. Introduction

From Experiments 1 and 2 it appeared that variations in a priori probability alone did not induce a strong response bias in three-category tasks. In addition, it appeared that observers' natural bias against equality responses, coupled with their assessment of the probabilities in the experiment, may have acted to obscure the already weak effects of the manipulation of a priori probability, especially when equal stimuli were highly probable. In order to obtain a clearer picture of the biassing effects of a priori probability, measure of accuracy, time and confidence were examined in a two-category task. At the same time, since there has been no systematic examination of the interaction between bias and discriminability, stimulus difference was varied within each level of bias. A between-subjects design allowed for the preliminary investigation of a broad range of bias conditions.

2. Method

(a) Stimuli

The stimuli were white pi-shaped figures presented against the medium grey background of the video monitor as described in Chapter 3, section B.2(a). As in Experiment 2, the horizontal line was of constant length, 80 mm, and always appeared in the same position on the screen. The vertical lines always varied in length and the difference in their lengths defined the discriminability of the stimulus. The longer vertical line, which could occur on either the right or the left side, was always 187 mm in length while the shorter line varied between 186.5 mm and 184.5 mm in steps of 0.5 mm. Thus, there were five levels of discriminability which were labelled from 1, the smallest stimulus difference, to 5, the largest stimulus difference. These differences in length subtended a visual angle of less than 0.1° when viewed at a distance of four metres.

(b) Apparatus

The apparatus was identical to that used in Experiment 2 and described in Chapter 3, section B.2(b) except that the middle response button was not present.

(c) Observers

The 50 observers, 25 men and 25 women, were students enrolled in the first year psychology course at the University of Adelaide, whose participation was credited towards a course requirement. They ranged in age from 17 to 24 years and all declared themselves to be right-handed. All observers were naive with respect to the aims of the experiment.

(d) Design

Five groups of 10 observers were formed, each comprising five men and five women. The groups performed under different experimental conditions distinguished by the probability that the right line of the stimulus was longer. The probabilities used for the five groups were 0.1, 0.3, 0.5, 0.7 and 0.9, and these groups were labelled Condition 1 to Condition 5, respectively. Observers were tested individually for approximately one hour.

For all conditions, stimuli were presented in 9 blocks: 4 practice blocks each of 50 trials preceded 5 experimental blocks each of 100 trials. Within each block, both practice and experimental, the occurrence of the longer line on the right or the left was random, given the probability restriction.

The five levels of discriminability appeared equally often on a given side in any condition and in a random sequence. Each level occurred on one fifth of the trials for which the given side was the longer.

Blocks of practice trials were separated by short rest periods in which observers were given the opportunity to clarify instructions. Consecutive blocks of experimental trials were separated by a 5 sec pause.

(e) Procedure

The procedure was identical to that used in Experiment 2 and described in Chapter 3, section B.2(e) with one exception. In this experiment, observers were instructed that the vertical lines would never be equal in length and that they should always choose the longer of the vertical lines. Observers were told to press the left button when the left line was chosen as the longer, and the right button when the right line was chosen as the longer. Observers were not told the stimulus probability restriction of the condition in which they were performing. Confidence ratings were obtained in exactly the same way as in Experiment 2, as described in Chapter 3, section B.2(e).

3. Results

Analyses of variance with Condition (proportion of right longer stimuli, 0.1, 0.3, 0.5, 0.7 or 0.9), Block (five levels, each block comprising 100 trials), Discriminability (five levels as defined in section A.2(a)) and Response (right or left), as factors were used to analyse data from the 500 experimental trials. Analyses of variance were carried out according to a crossed factorial design with repeated measures on Block, Discriminability and Response. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses. For the analyses of response time and confidence data, missing values were replaced with the overall mean value for that observer. This substitution was necessary because the majority of missing values occurred in Conditions 1 and 5 where the probabilities of right stimuli were 0.1 and 0.9, respectively, and when those cases which included missing values were omitted, no observers remained in these two groups.

(a) Accuracy

The analysis of accuracy data showed significant main effects due to Discriminability (F(4,180) = 340.15, MSerror = 0.16, $p < .001$) and Block (F(4,180) = 3.63, MSerror = 0.15, $p < .01$). In addition, there were significant interactions of Condition x Response (F(4,45) = 8.75, MSerror = 1.52, $p < .001$), Condition x Response x Discriminability (F(16,180) = 16.05, MSerror = 0.25, $p < .001$) and Condition x Response x Discriminability x Block (F(64,720) = 1.52, MSerror = 0.14, $p < .01$).

The Discriminability main effect arose because the percentage of stimuli for which responses were correct increased as stimulus difference increased, but at a decreasing rate. The percentages of correct responses at Discriminability levels 1 to 5 were 73.87%, 87.39%, 94.15%, 96.79% and 97.46%, respectively.

The Block main effect was significant because accuracy tended to decrease during the experimental session. However, none of the differences in accuracy between blocks was significant. The percentages of stimuli for which responses were correct in Blocks 1 to 5 were 90.73%, 90.30%, 90.50%, 88.56% and 89.57%, respectively.

The interaction between Condition and Response was significant because responses to the less probable right stimuli in Conditions 1 and 2 were significantly more accurate than responses to the more probable left stimuli in these two conditions. Although the difference in accuracy for right and left responses was not significant in any other condition, responses to the less probable stimulus were more accurate in all conditions in which the stimuli were not equiprobable. In Condition 3, where the stimuli were equiprobable, responses to right stimuli were more accurate than responses to left stimuli. From figure 21 it appears that, in Conditions 1 and 5 where the less probable stimulus had a probability of 0.1, accuracy for the corresponding response was markedly lower than in Conditions 2 and 4 where the less probable stimulus had a probability of 0.3. However, these differences were not significant.

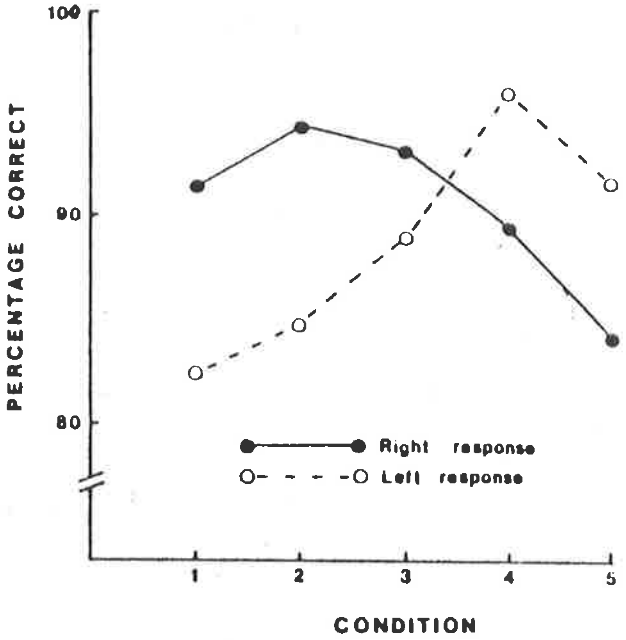


FIGURE 21. The percentage of right stimuli (solid line) and left stimuli (broken line) for which responses were correct in each condition. Percentages were taken over the 10 observers in each condition.

The interaction between Condition, Response and Discriminability arose because, at the two smallest levels of stimulus difference in all conditions in which the stimuli were not equiprobable (Conditions 1, 2, 4 and 5), accuracy tended to be much higher for responses to the less probable stimulus than for responses to the more probable stimulus. At the same time, in these conditions, accuracy for the two responses was equivalent at the three easier levels of discriminability. This is illustrated in figures 22(a) and 22(b) which show accuracy in responses to right and left stimuli at each level of discriminability for Condition 1 and Condition 5, respectively. Figure 22(c) shows the difference in accuracy for responses to right stimuli and accuracy for responses to left stimuli (right - left) at each level of discriminability, in each condition. From the figure it is apparent that, in Conditions 1, 2, 4 and 5, the size of the difference in accuracy decreased as stimulus difference increased. In Conditions 1 and 2, where right stimuli were less probable, the differences tended to be positive i.e. responses to right stimuli were more accurate than responses to left stimuli. In Conditions 4 and 5, where left stimuli were less probable, the differences tended to be negative i.e. responses to left stimuli were more accurate than responses to right stimuli. In Condition 3, where the stimuli were equiprobable, the curve remained fairly flat across all levels of discriminability, indicating that responses to both stimuli tended to be made with the same relative accuracy, irrespective of the level of discriminability.

The Condition x Response x Discriminability x Block interaction appears to have several components. Firstly, as described above, when the stimuli were not equiprobable, responses to the less probable stimulus were more accurate than responses to the more probable stimulus. Secondly, this finding was much more apparent at the smaller values of stimulus difference.

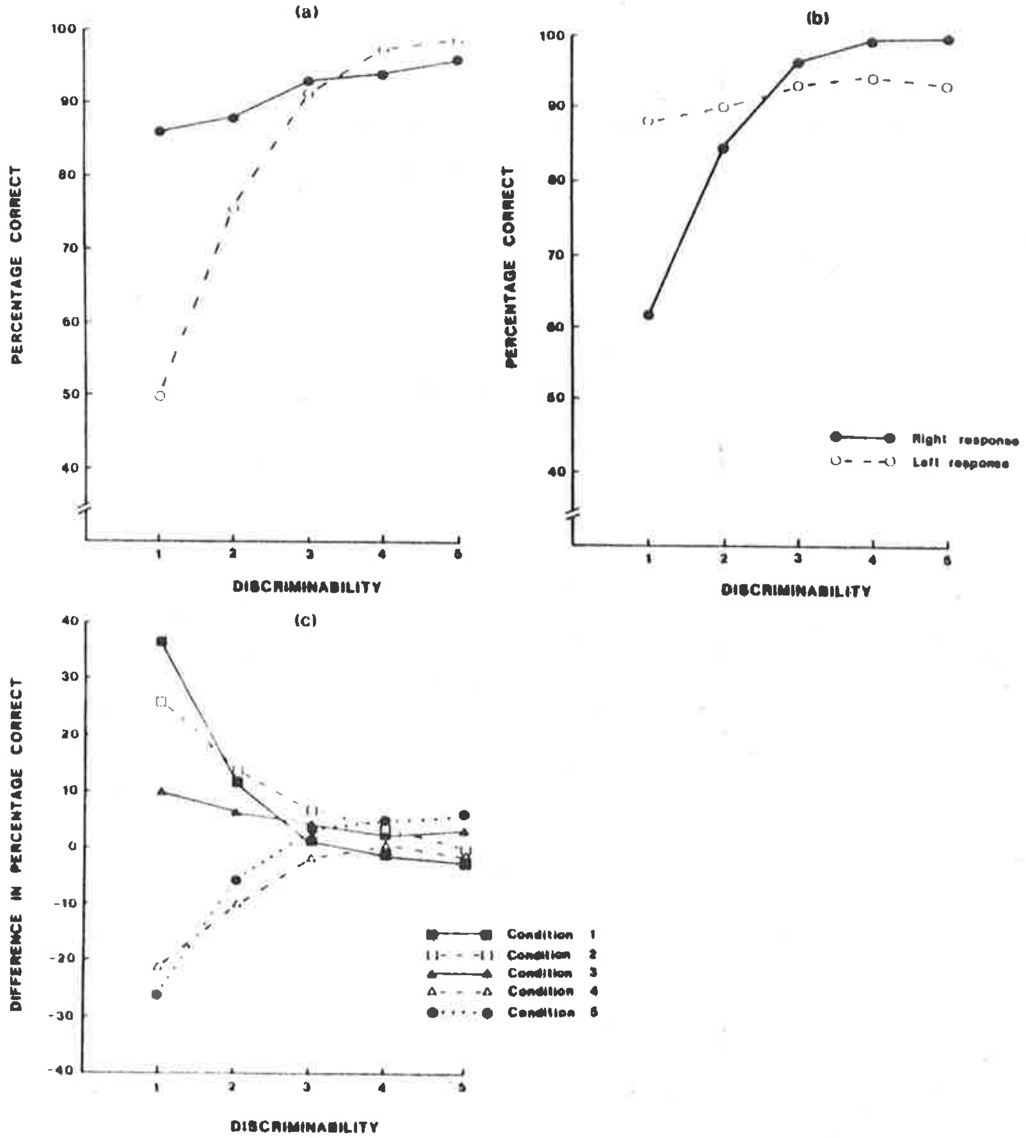


FIGURE 22. Figures 22(a) and 22(b) show the percentage of right stimuli (solid line) and left stimuli (broken line) for which responses were correct at each level of discriminability in (a) Condition 1 and (b) Condition 5, respectively. Figure 22(c) shows the difference in the percentage of correct responses to right stimuli and the percentage of correct responses to left stimuli, (right - left), at each level of discriminability in each condition. In all cases, the percentages were taken over the 10 observers in each condition.

This can be seen in figures 23(a) and 23(b) which show the difference in accuracy for the two responses (right - left) at each level of discriminability in each block, in Condition 1 and Condition 5, respectively. From the figures it is clear that, at Discriminability level 1 in Condition 1, the differences in each block were large and positive, indicating that responses to right stimuli were more accurate than responses to left stimuli. Conversely, at Discriminability level 1 in Condition 5, the differences in each block were large and negative showing that responses to left stimuli were more accurate than responses to right stimuli. The differences became smaller for the larger values of stimulus difference in both conditions, and the curve for one level of Discriminability, level 5 in Condition 1, included both positive and negative differences. Indeed, when data at the easier levels of discriminability were considered, there appeared to be some reversal of the finding that responses to the less probable stimulus were more accurate. This can be seen from figures 23(a) and 23(b). For example, in Condition 1, although all of the differences were positive at Discriminability level 1, at Discriminability level 5 the differences were negative in Blocks 3, 4 and 5 indicating that responses to left stimuli were more accurate than responses to right stimuli. In addition, in Block 1 there was no difference in accuracy between responses to right and left stimuli. Similarly, in Condition 5 at Discriminability level 1 all differences were negative, but at Discriminability level 5 all differences were positive. This indicates that responses to left stimuli changed from being more accurate than responses to right stimuli to being less accurate than responses to right stimuli.

Thirdly, in Condition 3, where stimuli were equiprobable, responses to right stimuli were always more accurate than responses to left stimuli.

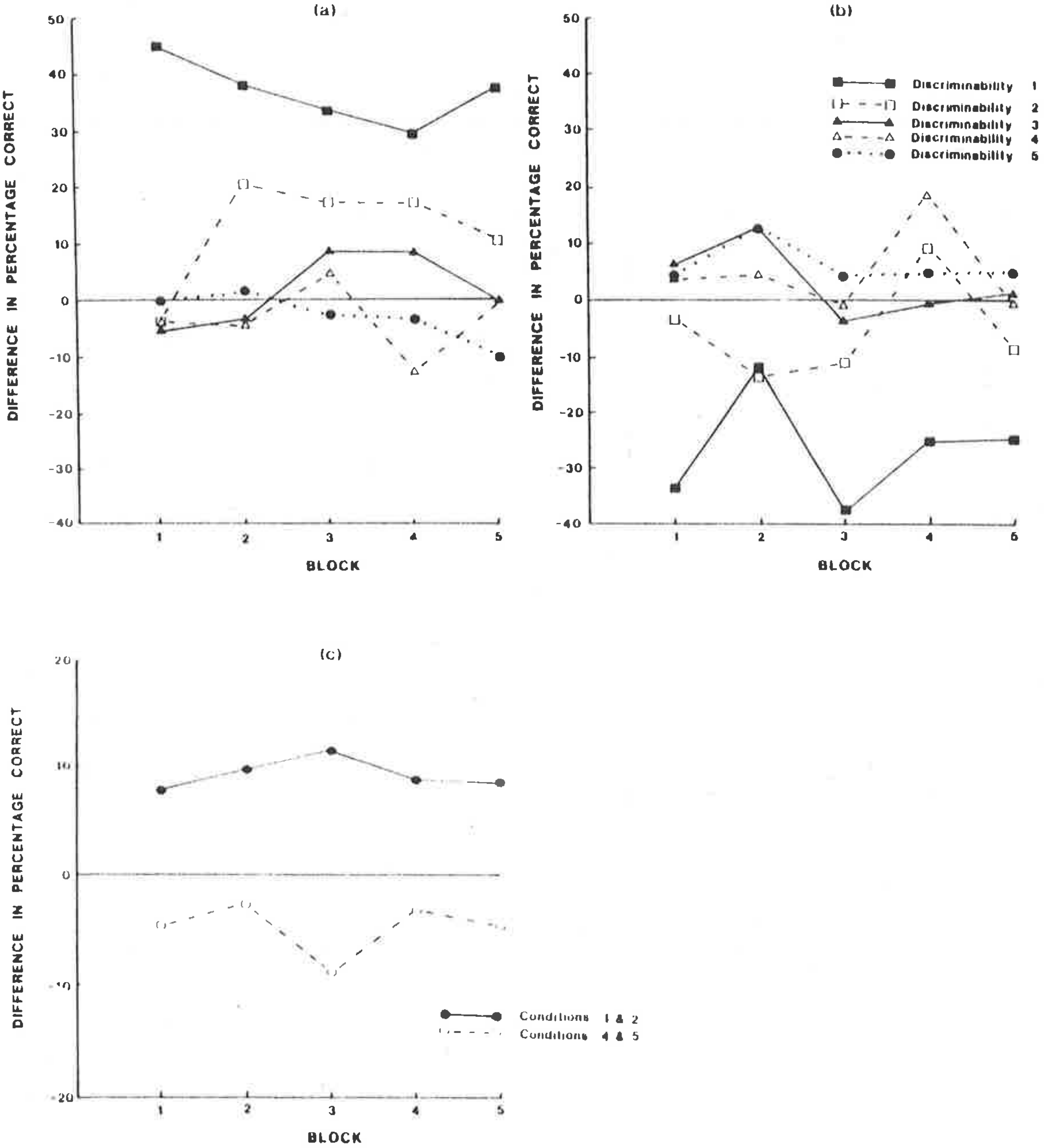


FIGURE 23. Figures 23(a) and 23(b) show the difference in the percentage of correct responses to right stimuli and the percentage of correct responses to left stimuli, (right - left), at each level of discriminability in each block in (a) Condition 1 and (b) Condition 5. Figure 23(c) shows the differences in the accuracy of responses to right and left stimuli in each block, (right - left), averaged over all discriminability levels in Conditions 1 and 2 (solid line) and all discriminability levels in Conditions 4 and 5 (broken line). In all cases, the differences were averaged over the 10 observers in each condition.

Fourthly, there was a change in the relative accuracy of the two responses across blocks of trials. Figure 23(c) shows the differences in accuracy for right and left responses in each block (right - left) averaged over all discriminability levels for Conditions 1 and 2 and all discriminability levels for Conditions 4 and 5. From the figure it appears that, both when right stimuli were more probable and when left stimuli were more probable, the size of the difference tended to increase from Block 1 to Block 3 and then decrease from Block 3 to Block 5.

(b) Response time

The analysis of response time data showed significant main effects due to Condition (F(4,45) = 2.99, MSerror = 360469, $p < .05$) and Discriminability (F(4,180) = 63.99, MSerror = 31664, $p < .001$) and a significant interaction between Condition and Response (F(4,45) = 4.14, MSerror = 264431, $p < .01$).

The Condition main effect was significant because, when the stimuli were equiprobable, response times were short overall, and they increased as the probabilities became more extreme. However, the increase was more marked in Conditions 1 and 2, where the left stimulus was the more probable, than in Conditions 4 and 5, where the right stimulus was the more probable. The overall mean response times for Conditions 1 to 5 were 822.14 msec, 787.91 msec, 701.80 msec, 772.83 msec and 803.59 msec, respectively.

The Discriminability main effect arose because response times decreased as stimulus difference increased, but at a decreasing rate. The mean times for responses at Discriminability levels 1 to 5 were 871.06 msec, 812.69 msec, 759.54 msec, 731.72 msec and 715.85 msec, respectively.

The interaction between Condition and Response is shown in figure 24 where it can be seen that, in Conditions 2, 3 and 4, right responses tended to

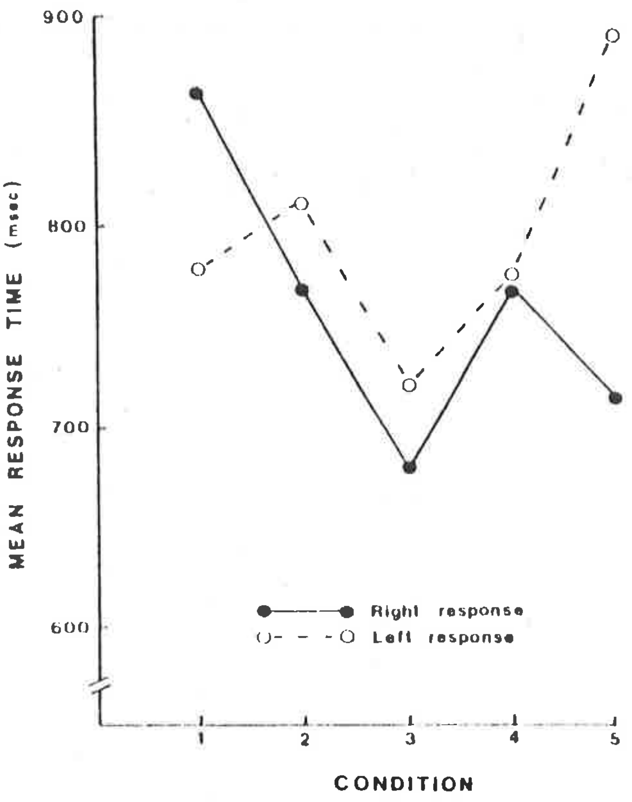


FIGURE 24. Mean times for correct right responses (solid line) and correct left responses (broken line) in each condition. The means were taken over the 10 observers in each condition.

be made more quickly than left responses, though the differences were not significant. In contrast, in Conditions 1 and 5, where the probabilities of the stimuli were most disparate, times for correct responses to the more probable stimulus were significantly shorter than times for correct responses to the less probable stimulus. In Condition 1, left responses were faster than right responses, while in Condition 5, right responses were faster than left responses. From the figure it is apparent that in Conditions 1 and 5, where the more probable stimulus had a probability of 0.9, times for this response were significantly shorter than times for the corresponding response in Conditions 2 and 4, where the more probable stimulus had a probability of 0.7. In Condition 3, where the stimuli were equiprobable, times for both right and left responses were significantly shorter than in any other condition.

(c) Confidence

The analysis of confidence data showed significant main effects due to Discriminability ($F(4,180) = 135.26$, MSerror = 0.51, $p < .001$) and Block ($F(4,180) = 5.37$, MSerror = 0.54, $p < .001$), and significant interactions of Condition x Response ($F(4,45) = 10.71$, MSerror = 2.25, $p < .001$) and Condition x Response x Discriminability ($F(16,180) = 5.04$, MSerror = 0.23, $p < .001$).

The main effect of Discriminability arose because confidence increased with stimulus difference, but at a decreasing rate. The mean confidence at Discriminability levels 1 to 5 was 2.50, 2.68, 3.01, 3.24 and 3.35, respectively.

The Block main effect reached significance because confidence in Block 1 was significantly higher than confidence in any other block. The other blocks did not differ significantly from one another. The mean confidence in Blocks 1 to 5 was 3.08, 2.93, 2.90, 2.89 and 2.92, respectively.

The Condition x Response interaction arose because confidence was higher for responses to the less probable stimulus than for responses to the more probable stimulus in Conditions 1, 2, 4 and 5 (see figure 25). In other words, confidence was higher for right responses in Conditions 1 and 2 and for left responses in Conditions 4 and 5. When stimuli were equiprobable in Condition 3, confidence was higher for right responses.

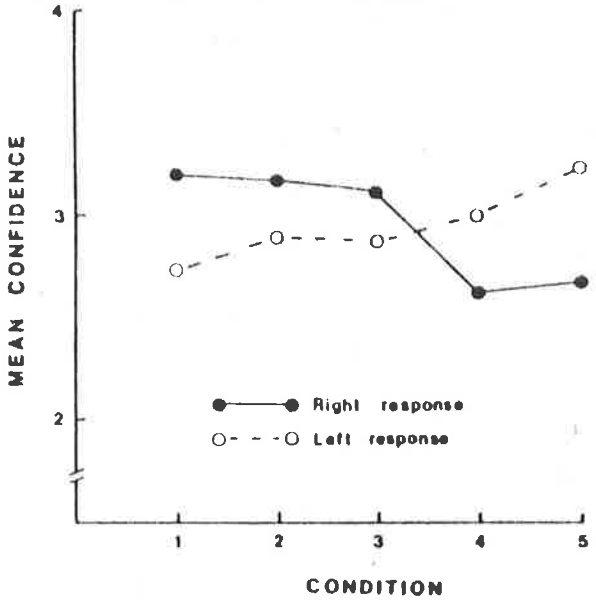


FIGURE 25. Mean confidence in correct right responses (solid line) and correct left responses (broken line) in each condition. The means were taken over the 10 observers in each condition.

The Condition x Response x Discriminability interaction is illustrated in figure 26. Figures 26(a) and 26(b) show the mean confidence in right and left responses at each level of discriminability in Condition 1 and Condition 5, respectively. From these figures it is apparent that, at all levels of discriminability in both of these conditions, confidence was higher in responses to the less probable stimulus. This was also the case in Conditions 2 and 4. In Condition 3, where the stimuli were equiprobable, confidence was higher in right responses. Figure 26(c) shows the difference in confidence in right and left responses (right - left) at each level of discriminability for each condition. The figure shows that in Conditions 1 and 2, where right stimuli were less probable, and in Condition 3, where the stimuli were equiprobable, all differences were positive i.e. confidence in right responses was higher than confidence in left responses. In Conditions 4 and 5 where right stimuli were more probable, all differences were negative i.e. confidence in right responses was lower than confidence in left responses. However, in all cases, it is apparent that the difference in confidence for the two responses decreased as stimulus difference increased.

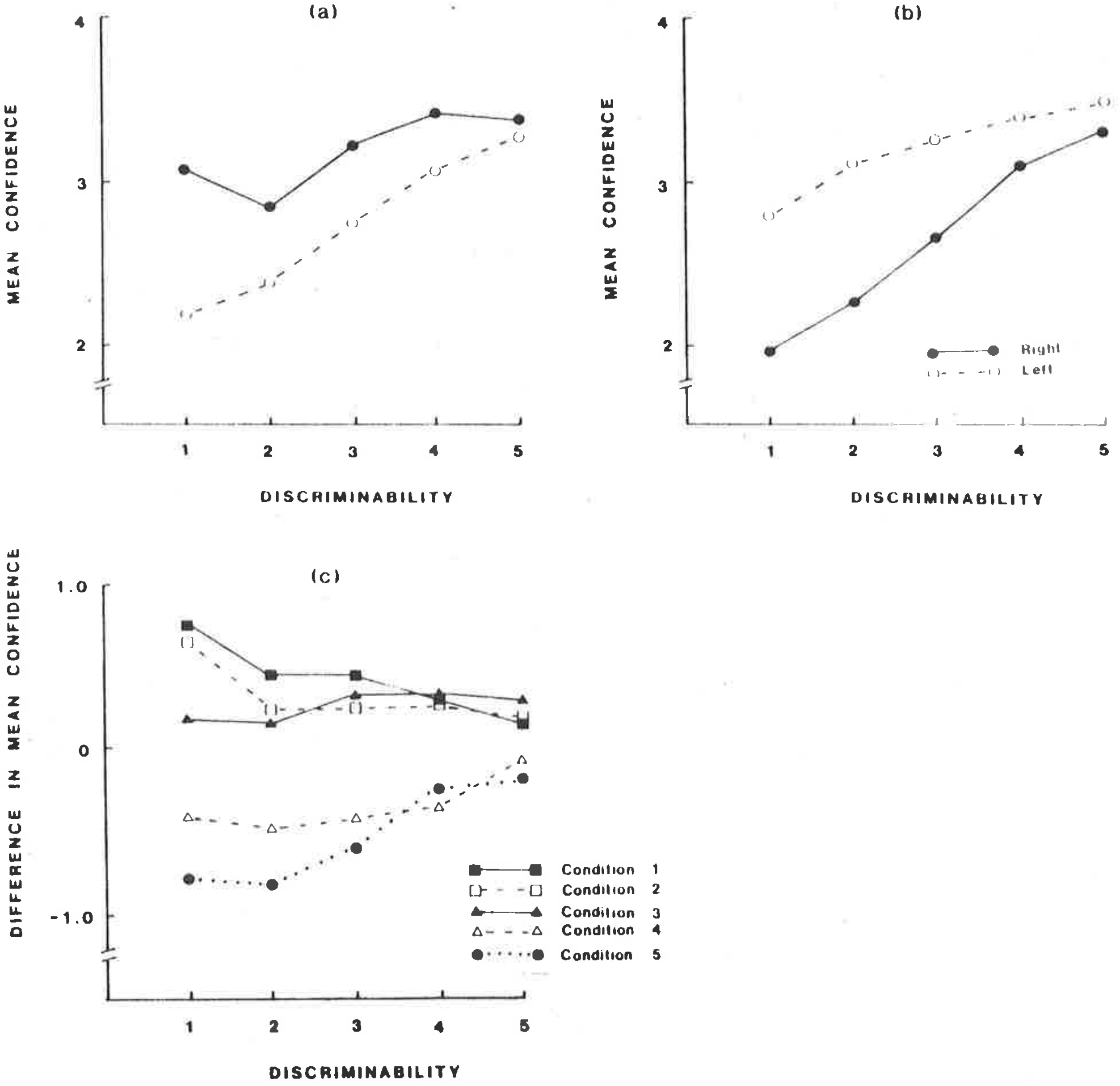


FIGURE 26. Figures 26(a) and 26(b) show the mean confidence in correct right responses (solid line) and correct left responses (broken line) at each level of discriminability in (a) Condition 1 and (b) Condition 5, respectively. Figure 25(c) shows the difference in confidence in correct right and left responses (right - left), at each level of discriminability in each condition. In all cases, the means were taken over the 10 observers in each condition.

(d) Confidence and response time

For each observer in each condition, confidence ratings were plotted against the corresponding response time values, and straight lines were fitted. This was done for right and left responses separately at each level of discriminability, producing 500 plots. It was found that 358 of the slopes of the best fitting straight lines were negative, showing that, overall, there was an inverse relationship between confidence and response time ($p < .001$, binomial test). This inverse relationship was still evident when right and left responses were considered separately for each condition. This can be seen from table 9 where the number of negative slopes of the 50 best fitting straight lines in each case is given. In all cases except right responses in Condition 1, the majority of best fitting straight lines had a negative slope.

CONDITION	NEGATIVE SLOPES RIGHT RESPONSES	p	NEGATIVE SLOPES LEFT RESPONSES	P
1	25	n.s.	44	<.001
2	40	<.001	48	<.001
3	40	<.001	41	<.001
4	43	<.001	39	<.001
5	44	<.001	37	<.001

TABLE 9. The number of negative slopes obtained when, for each observer in each condition, confidence ratings were plotted against the corresponding response time values and straight lines were fitted. Separate plots were made for data for right responses and for left responses at each level of discriminability, resulting in 50 plots for each type of response in each condition. The total number of negative slopes obtained in each case is shown with its associated binomial probability.

Little can be concluded from the discrepant results in Condition 1 since the probability of a right stimulus in this condition was 0.1, so the results being considered here were based on very few responses. This would also account for the less significant result for left responses in Condition 5. It appears that, independent of variations in stimulus probability, there was a predominantly inverse relationship between confidence and response time.

4. Discussion

(a) Accuracy

The direct relationship between accuracy and stimulus difference is a common finding in discrimination tasks (e.g. Henmon, 1906; Lemmon, 1927; Kellogg, 1931; Johnson, 1939; Festinger, 1943a,b) and is a simple prediction of each of the mechanisms for criterion regulation applied to any of the models, as described in Chapter 2.

The slight decrease in accuracy during the experimental session does not appear to indicate response bias but appears to be best explained in terms of a decrement in performance due to fatigue. This explanation is consistent with all of the mechanisms, though the effect is not specifically predicted by them. However, since the changes in accuracy which gave rise to this effect were very small (<2%) it must be interpreted with reservation.

The finding that responses to the less probable stimulus were more accurate than responses to the more probable stimulus contrasts with Laming's (1968, 1969) findings and cannot be explained in terms of some of the mechanisms for criterion regulation. For example, according to the ideal observer hypothesis applied to Signal Detection Theory, when one stimulus is highly probable, the cutoff will move to a position where observations will be more likely to fall on the side of the cutoff for the corresponding response. Therefore, the response which is appropriate to the more probable stimulus will be made more often than the response which is appropriate to the less probable stimulus. As a consequence, it seems likely that errors will take the form of responses appropriate to the more probable stimulus being made when the less probable stimulus is presented. In other words, the ideal observer hypothesis in Signal Detection Theory predicts a pattern of results which is the opposite of the one found in the present experiment. This is also the case when the ideal observer hypothesis is combined with the random walk or accumulator models.

These data are consistent with the response stabilisation hypothesis of Vickers et al. (1977) in combination with any of the models if it is assumed that the observer's initial assessment of the stimulus probabilities tends to be less extreme than the actual probabilities. For example, as in this experiment, when observers are not given any information about actual stimulus probabilities and stimuli are equiprobable in the practice trials, it seems likely that their original expectation will be that the stimuli will continue to be equiprobable. In this situation, the local probability of responses to the more probable stimulus will tend to exceed the cumulative probability, or the observer's expectation of the probability of this responses and the corresponding criterion will be adjusted upwards. As a consequence, accuracy of responses to the more probable stimulus will decrease.

These data are also consistent with the adaptation-level hypothesis of Vickers and Leary (1983). In this case, the mean of the sensory intensities of all observations will tend to take a value close to the intensity of the more probable stimulus and so, in Signal Detection Theory terms, the cutoff will tend to move in the direction of the more probable stimulus. Alternatively, in a random walk process or the accumulator model, the indifference point will take a value closer to the intensity of the more probable stimulus so that the effective discriminability of the more probable stimulus is reduced. As a consequence, in all cases, fewer responses which are appropriate to the more probable stimulus will be made and more responses which are appropriate to the less probable stimulus will be made. Therefore, accuracy will be higher in responses to the less probable stimulus.

The current data are not consistent with the target confidence mechanism according to which unequal a priori stimulus probabilities will give rise to changes in the relative values of the criteria, so that the criterion for responses to the more probable stimulus will be low relative to the criterion for responses to the less probable stimulus. Therefore, the response appropriate to the more probable stimulus will have a high probability of occurrence, both when it is correct and when it is in error, so that accuracy will be higher for these responses.

None of the mechanisms for criterion regulation predicts the drop in the accuracy of responses to the less probable stimulus in Conditions 1 and 5. However, this change in accuracy could be accommodated in terms of a shift in criterion values, if it is assumed that, in these conditions, the observers became aware of the large discrepancy in the accuracy rates for right and left responses and tried to compensate for it. Further evidence for this interpretation is discussed below in terms of the interaction of Condition x Response x Discriminability x Block.

When the Condition x Response interaction is considered in the case where the stimuli were equiprobable (Condition 3), responses to right stimuli were more accurate than responses to left stimuli. This appears to indicate that observers had a pre-existing bias towards right responses independent of the manipulation of a priori probability. This is not surprising since all observers were right-handed. Although this result is not predicted by any of the mechanisms, it is not inconsistent with any of them.

The actual changes in accuracy which gave rise to the Condition x Response x Discriminability interaction are predicted by the response stabilisation and adaptation-level hypotheses, as described above. However, the general finding that the Condition x Response effect applied at each level of Discriminability is consistent with all of the mechanisms. In addition, it appears that the effects of a priori probability were shown more strongly at the more difficult levels of discriminability, where the differences in accuracy were greatest. As with the Condition x Response interaction, this becomes more meaningful when considered with reference to the Condition x Response x Discriminability x Block interaction.

Each component of the Condition x Response x Discriminability x Block interaction will be discussed in turn. As seen above, the finding that when the stimuli were not equiprobable, responses to the less probable stimulus were more accurate, is consistent with the response stabilisation hypothesis and the adaptation-level hypothesis, while the finding that right responses were more accurate than left responses in Condition 3, where the stimuli were equiprobable, appears to arise from a bias to right responses of right-handed observers.

In addition, in these data, the finding that responses to the less probable stimulus were more accurate was more apparent at lower levels of discriminability. Although this is not a specific prediction of any of the mechanisms, it is consistent with an explanation in terms of variable criteria. However, the suggestion that the trend reversed, so that responses to the more probable stimulus were more accurate at higher levels of discriminability, contrasts with the predictions of all of the mechanisms. In order to explain this finding it would be necessary to assume that observers adopted different criterion levels when responding to stimuli of different levels of difficulty. Since the different levels of discriminability were presented in a random order, this explanation cannot be accommodated by any of the mechanisms.

On the other hand, one possible explanation for this trend was suggested by the evidence in Condition 3 for a bias for right responses independent of _a priori_ probability. The tendency for responses to the more probable stimulus to be more accurate at Discriminability level 5, which was most marked in Condition 5, may arise from the bias towards right responses. It seems likely that this bias would be more evident at easy levels of discriminability where it is clear which line is longer, and more especially when right stimuli were the more probable, as in Condition 5. This finding would not be inconsistent with any of the mechanisms.

Finally, there appeared to be some variation in the relative accuracy of responses to right and left stimuli across blocks. The size of the difference in accuracy tended to increase from Block 1 to Block 3 and then decrease from Block 3 to Block 5. In order to explain this in terms of variable criteria, it is necessary to assume that the criterion values for right and left responses were similar in Block 1, but that they were adjusted so that one was much higher than the other in Block 3 and then they returned to take similar values in Block 5. This form of criterion adjustment is not a simple prediction of any of the mechanisms for criterion regulation. However, as suggested by Vickers (1979), it is possible that criterion values may be influenced by the expectations of the observer which may not correspond to the actual stimulus situation. For example, in this experiment it seems likely that, since observers were not told of the differences in the probabilities of the stimuli, they would have made an assessment of the probabilities during the practice trials. As there had been no indication that probabilities might be very disparate, it is possible that the observers' assessments were not as extreme as the actual probabilities. Therefore, to begin with, their criteria for the two responses would be relatively similar. Consequently, in Condition 1 for example, as the session progressed observers would find that they were making more left responses than was consistent with their assessment of the probabilities. In order to compensate for this mismatch, observers might increase the criterion for left responses relative to the criterion for right responses. With the progress of trials, it is likely that the observers' assessments of the probabilities would become more accurate, resulting in another change in criteria, and leading to a more appropriate representation of the actual stimulus probabilities. However, it appears that this adaptation was not complete by the end of Block 5.

If this explanation applies, the finding that responses to the less probable stimulus were more accurate is no longer inconsistent with any of the mechanisms for criterion regulation in combination with any of the decision models. This result can be viewed as the product of an inappropriate compensatory shift in criteria by observers who had underestimated the difference in the probabilities of the two stimuli.

(b) Response time

The finding that response times were shortest in Condition 3 and increased as the stimulus probabilities became more extreme appears to have two components. Firstly, the shortest time in Condition 3 seems to indicate that observers found this condition, in which the stimuli were equiprobable, less difficult. This may be because, in this condition alone, the actual stimulus situation agreed with the observers' expectations based on the experience of the practice trials, that the stimuli would be equiprobable. Secondly, the more marked increase in Conditions 1 and 2 than in Conditions 4 and 5 may be interpreted in terms of a bias towards right responses arising from the right-handedness of observers. In Conditions 1 and 2, where the majority of stimuli were left stimuli, most responses would have been the slower left responses. Therefore, the overall mean response time would be longer than in Conditions 4 and 5 where the majority of responses to highly probable right stimuli would have been the faster right responses. Although this finding can be accommodated by all of the mechanisms, it does not correspond to a specific prediction of any of them.

The inverse relationship between response time and discriminability is a well established finding in the literature (e.g. Henmon, 1906; Lemmon, 1927; Kellogg, 1931; Johnson, 1939; Festinger, 1943a,b) and is predicted by all of the mechanisms.

The Condition x Response interaction showed that, in Conditions 1, 2, 4 and 5, responses to the more probable stimulus were made more quickly than responses to the less probable stimulus, while in Condition 3, left responses were made more slowly than right responses. These differences in times are consistent with Laming's (1968, 1969) results and can be explained by the ideal observer hypothesis applied to Signal Detection Theory, by the choice of an appropriate latency function for each condition. Alternatively, when the ideal observer hypothesis is applied to a random walk or an accumulator model, the criterion amount of information necessary to make the response to the more probable stimulus will be low, so that times for this response will be short, as seen in these data. The pattern of results in Conditions 1 and 5 is also consistent with the target confidence mechanism applied to either a random walk or an accumulator model. In either case, the criterion for responses to the more probable stimulus will be low relative to the criterion for responses to the less probable stimulus and, as a result, it is likely to be satisfied more quickly. In this way, responses to the more probable stimulus are likely to be faster than responses to the less probable stimulus, as is seen in these data in Conditions 1 and 5. The results of Conditions 2, 3 and 4 are not as simply accommodated by the target confidence mechanism. In particular, the different patterns shown by Conditions 1 and 2 contrast with the predictions of the target confidence mechanism. However, it seems likely that the right-handedness of the observers is producing the shorter times for right responses, at least in Condition 3 and possibly in Conditions 2 and 4 as well.

These data conflict with the response stabilisation and adaptation-level hypotheses. As seen above, according to the response stabilisation hypothesis, the criterion for responses to the more probable stimulus will be high relative to the criterion for responses to the less probable stimulus, while according to the adaptation-level hypothesis, the effective discriminability of the more probable stimulus will be reduced. Therefore, in both cases more observations will be needed before a response can be made and so responses will be slower.

(c) Confidence

The direct relationship between confidence and stimulus difference is a common finding in previous studies (e.g. Garrett, 1922; Johnson, 1939; Festinger, 1943a; Pierrel & Murray, 1963) and is consistent with all of the mechanisms.

The finding that confidence decreased from Block 1 to Block 2 appears to provide further evidence for Vickers' (1979) notion of the influence of expectation. If, as suggested above, at this point observers were becoming aware that their assessment of the stimulus probabilities did not correspond to their pattern of responding, they may have increased the criterion for responses to the more probable stimulus which they appeared to be making too often. This variation in criteria is consistent with all of the mechanisms. As a result, the response to the less probable stimulus would increase in frequency, and since the criterion for this response is low, according to the balance of evidence hypothesis, confidence will be low. On the other hand, these data are not consistent with Audley's (1960) hypothesis. When response time and confidence data are considered together, there is a direct relationship between confidence and time in Conditions 1, 2 and 3.

When considered in the absence of response time data, the finding that confidence was higher in responses to the less probable stimulus is predicted by the response stabilisation and adaptation-level hypotheses applied to Signal Detection Theory or the random walk model. These mechanisms predict that times for responses to the less probable stimulus will be short so, following Audley (1960), confidence will be high. However, when response time and confidence data are considered together, the Condition x Response interaction is not consistent with Audley's (1960) hypothesis in combination with any of the criterion regulation mechanisms applied to Signal Detection Theory or the random walk model. In Conditions 4 and 5, left responses were made more slowly but with higher confidence than right responses. However, when the response stabilisation hypothesis is considered with the balance of evidence hypothesis it would be predicted that confidence will be higher for responses to the more probable stimulus. This is opposite to the trend shown in these data.

On the other hand, according to the balance of evidence hypothesis of the accumulator model in combination with the mechanism based on target confidence, confidence should be higher for responses to the less probable stimulus because the criterion for this response will tend to be high relative to the criterion for responses to the more probable stimulus. Therefore, the maximum difference between the accumulated totals is possible when the higher criterion is satisfied, and consequently, confidence will tend to be higher for this response. This is the pattern shown in Conditions 1, 2, 4 and 5. However, when the finding in Conditions 1, 2, 4 and 5 is considered together with response time data there is some inconsistency. The shorter times for responses to the less probable stimulus imply that the criterion for this response is low and therefore confidence should be low. In addition, this explanation does not include the case of Condition 3 where the stimuli were equiprobable. According to the target confidence mechanism, in this situation, confidence should be the same for both responses. From figure 26 it is apparent that this was not the case in these data. In Condition 3, confidence was higher for right responses than for left responses.

This result is also inconsistent with any explanation in terms of the bias towards right responses of right-handed observers. If bias towards right responses is accommodated by supposing that the observer had a relatively low criterion for this response, it would be predicted that right responses would be less rather than more confident than left responses.

The finding that responses to the less probable stimulus tended to be more confident was also obtained at each level of discriminability giving rise to the Condition x Response x Discriminability interaction. Although this finding, considered on its own, is consistent with the balance of evidence hypothesis, when it is considered in conjunction with response time data it is not consistent with any of the processes.

(d) Confidence and response time

When the relationship between confidence and response time was considered within bias conditions, there was overwhelming evidence that confidence varied inversely with response time. As seen above, this relationship is consistent with Audley's (1960) hypothesis and the balance of evidence hypothesis and may therefore be accommodated by any of the models. However, this conclusion must be presented with caution in the light of the conflicting results when variations in time and confidence are considered between bias conditions.

5. Summary and Conclusions

When the effects of a priori probability on the three dependent variables are considered together, all of the changes observed were not consistent with any one of the mechanisms for criterion control in the models. However, the influence of discriminability was systematic.

As in Experiment 1, accuracy and confidence increased and response time decreased as stimulus difference increased. This finding is predicted by all of the mechanisms.

The effects of the manipulation of a priori probability were evident in accuracy and confidence measures, while response time data remained unclear. In all conditions in which stimuli were not equiprobable, responses to the less probable stimulus were more accurate and were made with higher confidence. However, right responses tended to be made more quickly than left responses at each level of discriminability in each condition.

The findings for accuracy and confidence do not correspond to a simple prediction of any of the mechanisms, although when considered singly, each set of data can be explained in terms of a different mechanism. Accuracy data are consistent with the response stabilisation or adaptation-level hypotheses in combination with any of the models, while the pattern shown by confidence is predicted by the balance of evidence hypothesis in combination with the target confidence mechanism. At the same time, the finding that left responses were slower than right responses is consistent with any of the models but when it is considered with confidence data, problems arise for all of the mechanisms.

In addition, when the relationship between confidence and response time was considered within bias conditions, as predicted by both Audley's (1960) hypothesis and the balance of evidence hypothesis, confidence was higher when responses were faster. On the other hand, when these measures were examined between bias conditions there was no consistent variation and, in some cases, the relationship between them was direct.

Overall, the findings from this experiment appear to be consistent with an explanation in terms of variable criteria. However, the influence of the observers' assessments of the stimulus probabilities appears to be an important determinant of criterion regulation in the situation where they are not told the actual stimulus probabilities. As a result, there appeared to be factors acting in this experimental situation which did not correspond to specific predictions of the models. In addition, in order to examine the differences in the response to different bias conditions, comparison had to be made between probability conditions which were performed by different groups of observers. As a consequence, such comparisons can be made only with some reservation.

B. EXPERIMENT 4

1. Introduction

In Experiment 3, the data from the three dependent variables considered together were not consistent with any one of the mechanisms of criterion regulation as applied to the three decision models. It appeared that the influence of both the observer's assessment of the stimulus probabilities and his right-handedness may have masked the effects of *a priori* probability. Since probability was varied according to a between-subjects design, comparisons of effects between different bias conditions were confounded with individual differences. In addition, in Experiment 3, the relationship between confidence and response time appeared to vary according to whether it was examined within bias conditions or between bias conditions. Therefore, in Experiment 4, the effects of *a priori* probability on accuracy, time and confidence were examined using the same task as in Experiment 3, but in a within-subjects design. As a result, the effects of changes in probability could be examined independently of the effects of individual differences, while the relationship between confidence and response time could be studied within and between bias conditions performed by the same observer.

2. Method

(a) Stimuli

The stimuli were identical to those used in Experiment 3 and described in section A.2(a).

(b) Apparatus

The apparatus was identical to that used in Experiment 3 and described in section A.2(b).

(c) Observers

The 20 observers, 10 men and 10 women, were students in the first year psychology course at the University of Adelaide, whose participation was credited towards a course requirement. They ranged in age from 17 to 26 years, and all declared themselves to be right-handed. All observers were naive with respect to the aims of the experiment.

(d) Design

Two groups of 10 observers were formed, each comprising five men and five women. The groups performed under different order conditions. Observers were tested individually for approximately one hour.

In both order conditions, stimuli were presented in 11 blocks: 4 practice blocks each of 50 trials preceded 7 experimental blocks each of 100 trials. For both groups of observers, in each practice block and the first experimental block, the occurrence of the longer line on the right or the left was equiprobable ($p = 0.5$). The remaining six blocks of experimental trials took one of two forms:

A. The probability of the longer line occurring on the right was 0.8 and therefore, the probability of the longer line occurring on the left was 0.2.

B. The probability of the longer line occurring on the right was 0.2 and therefore, the probability of the longer line occurring on the left was 0.8.

One group of observers, Order AB, performed the last six experimental blocks in order ABABAB, while the other group of observers, Order BA, performed the last six experimental blocks in order BABABA. For both groups of observers the first two of these experimental blocks were treated as practice and omitted from the analyses. Only the last four experimental blocks were considered in the analyses.

Within each block, both practice and experimental, the occurrence of the longer line on the right or the left was random, given the probability restriction. The five levels of discriminability appeared equally often on a given side in blocks of type A or blocks of type B, and in a random sequence. Each level of discriminability occurred on one fifth of the trials for which the given side was longer.

Practice blocks were separated by short rest periods in which observers were given the opportunity to clarify instructions. The seven blocks of experimental trials were performed in an uninterrupted sequence.

(e) Procedure

The procedure was identical to that of Experiment 3 and described in section A.2(e).

3. Results

Analyses of variance with Order (AB or BA, as defined in section B.2(d)), Period (first two experimental blocks or second two experimental blocks), Probability (block type A or type B, as defined in section B.2(d)), Half (first or second half of an experimental block), Discriminability (5 levels, as defined in section A.2(a)) and Response (right or left) as factors were used to analyse data from the last 4 experimental blocks. Analyses of variance were carried out according to a crossed factorial design with repeated measures on Period, Probability, Half, Discriminability and Response. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses.

(a) Accuracy

The analysis of accuracy data showed a significant main effect of Discriminability ($F(4,72) = 70.12$, $MSerror = 0.38$, $p < .001$) and significant interactions of Probability x Half x Response ($F(1,18) = 33.80$, $MSerror = 0.18$, $p < .001$), Period x Probability x Half x Discriminability ($F(4,72) = 3.38$, $MSerror = 0.32$, $p < .05$), Period x Half x Discriminability x Response ($F(4,72) = 3.80$, $MSerror = 0.34$, $p < .05$) and Probability x Half x Discriminability x Response ($F(4,72) = 3.00$, $MSerror = 0.29$, $p < .05$).

The Discriminability main effect reached significance because accuracy increased as a direct function of stimulus difference, but at a decreasing rate. The percentage of stimuli for which responses were correct at Discriminability levels 1 to 5 were 76.37%, 87.42%, 93.44%, 95.39% and 96.95%, respectively.

The interaction between Probability, Half and Response is shown in figure 27 where it can be seen that the pattern of change during a block of trials depended on the probability of a response. In both types of block, accuracy of responses to the more probable stimulus decreased from the first to the second half of blocks, though the differences were not significant. In contrast, accuracy of responses to the less probable stimulus increased significantly from the first to the second half of blocks. In addition, the figure shows that, in the first half of blocks, responses to the more probable stimulus were more accurate than responses to the less probable stimulus. In this way, the respective decrease and increase in accuracy during the block of trials resulted in very similar rates of accuracy for both responses by the second half of the block of trials. From the figure it can also be seen that accuracy of responses to right stimuli in Half 2 of blocks of type A was significantly higher than accuracy of responses to right stimuli in Half 1 of blocks of type B. Accuracy of responses to left stimuli decreased at the same time, but the difference was not significant. Similarly, there was a significant increase in the accuracy of responses to right stimuli from Half 2 of blocks of type B to Half 1 of blocks of type A, while the decrease in accuracy of responses to left stimuli across this block boundary was not significant.

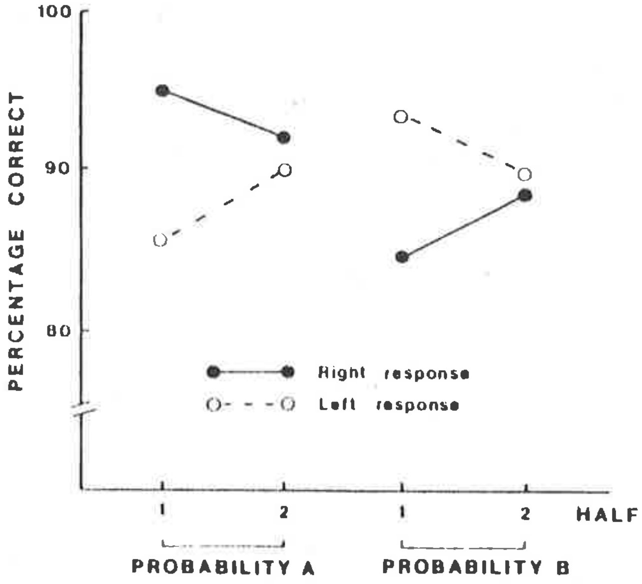


FIGURE 27. The percentage of right stimuli (solid line) and left stimuli (broken line) for which responses were correct in Half 1 and Half 2 of blocks of Probability A (left pair of lines) and blocks of Probability B (right pair of lines). The percentages were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

The interaction between Period, Probability, Half and Discriminability and the interaction between Period, Half, Discriminability and Response can be seen in figures 28 and 29, respectively. It appears that both of these interactions reached significance because of apparently unsystematic fluctuations in accuracy rates between periods. This random variation was most evident at the easiest levels of Discriminability, levels 3, 4 and 5. There was no clear pattern of change between periods in either set of data.

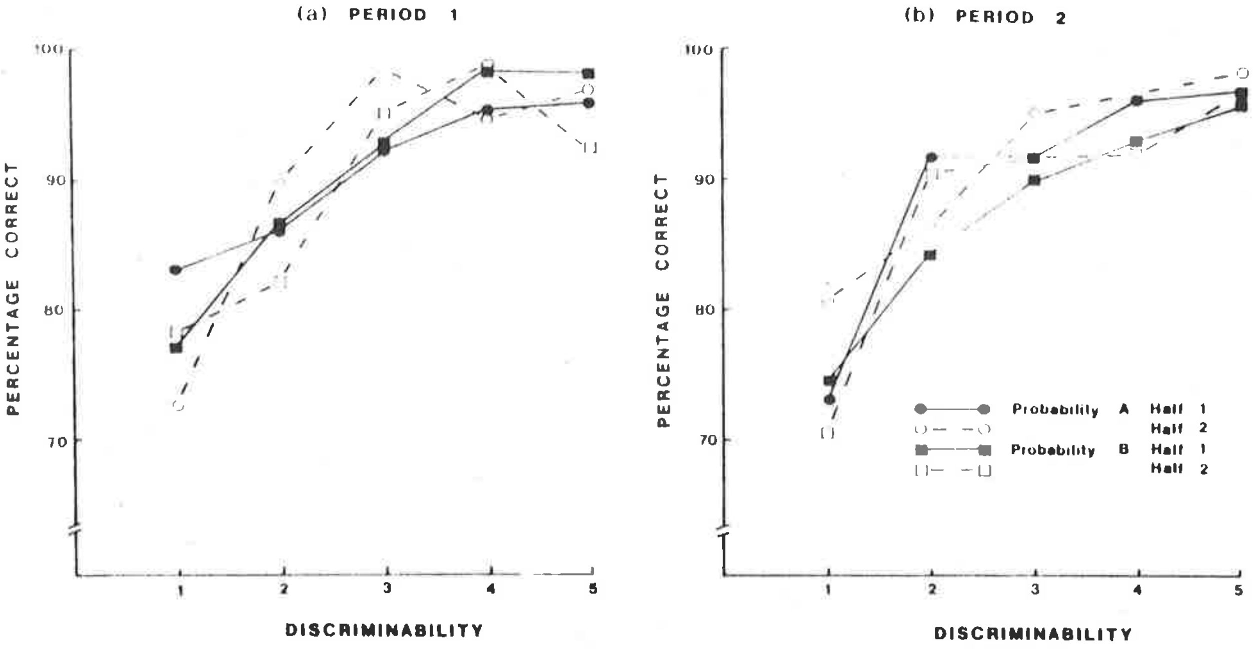


FIGURE 28. The percentage of stimuli for which responses were correct at each level of discriminability in Half 1 (solid line) and Half 2 (broken line) of blocks of Probability A (circles) and blocks of Probability B (squares) in (a) Period 1 and (b) Period 2. The percentages were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

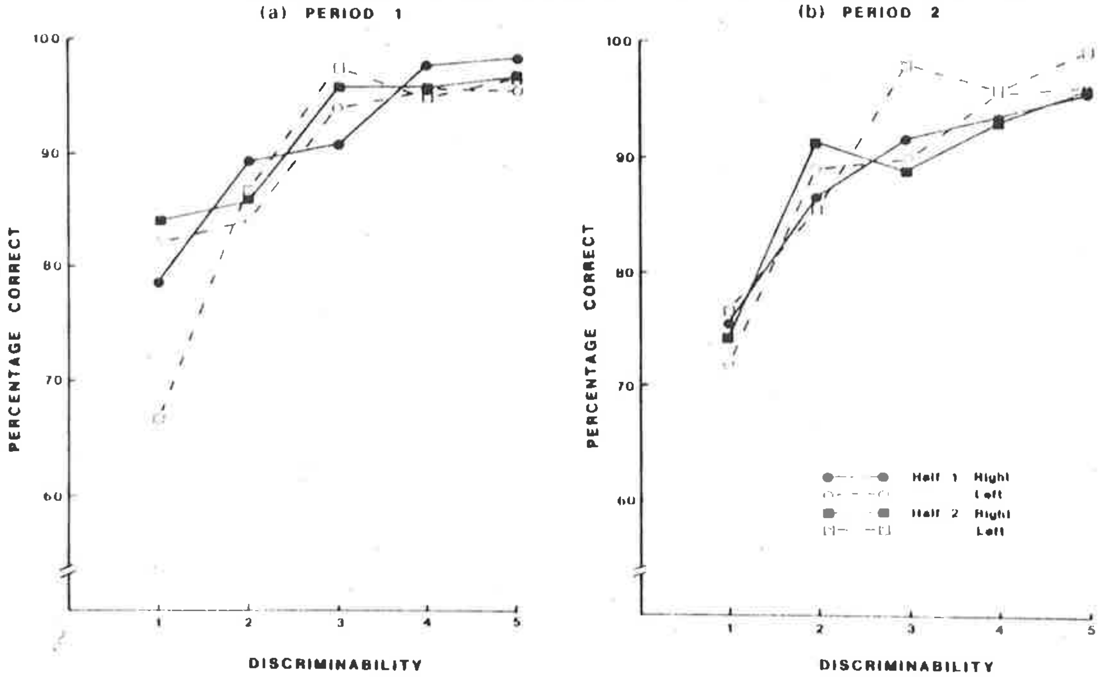


FIGURE 29. The percentage of right stimuli (solid line) and left stimuli (broken line) for which responses were correct at each level of discriminability in Half 1 (circles) and Half 2 (squares) of blocks in (a) Period 1 and (b) Period 2. The percentages were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

As shown in figure 30, the tendency for accuracy rates to become more similar from the first to the second half of blocks was still evident when data were considered for each level of discriminability, especially for blocks of type B. In addition, responses to right stimuli in the first half of blocks of type A tended to be more accurate than responses to right or left stimuli at any other time. This trend resulted in a significant interaction between Probability, Half, Discriminability and Response.

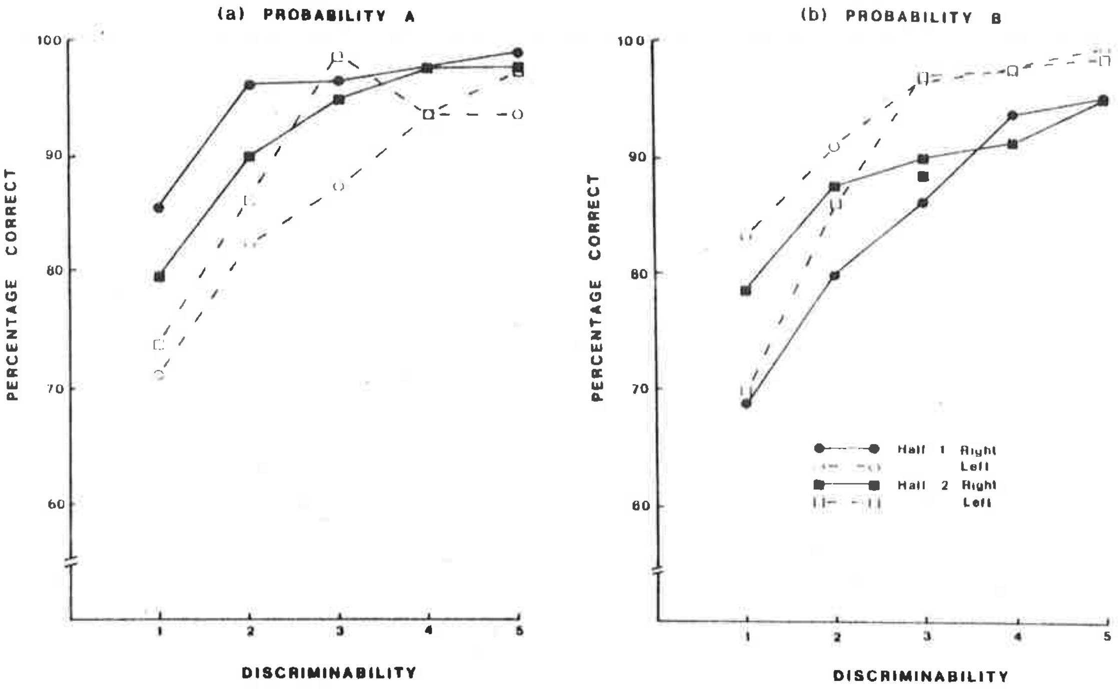


FIGURE 30. The percentage of right stimuli (solid line) and left stimuli (broken line) for which responses were correct at each level of discriminability in Half 1 (circles) and Half 2 (squares) of blocks of (a) Probability A and (b) Probability B. The percentages were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

(b) Response time

The analysis of response time data showed significant main effects of Half (F(1,18) = 6.68, MSerror = 24543, $p < .05$) and Discriminability (F(4,72) = 28.72, MSerror = 69325, $p < .001$) and significant interactions of Probability x Response (F(1,18) = 46.94, MSerror = 92989, $p < .001$), Probability x Half x Response (F(1,18) = 31.78, MSerror = 20294, $p < .001$), Period x Probability x Half x Discriminability (F(4,72) = 3.63, MSerror = 29859, $p < .01$), Period x Half x Discriminability x Response (F(4,72) = 3.67, MSerror = 38425, $p < .01$) and Probability x Half x Discriminability x Response x Order (F(4,72) = 3.16, MSerror = 30429, $p < .05$).

The Half main effect arose because response times were longer in the first half than in the second half of blocks. The mean times for responses in Half 1 and Half 2 of blocks were 802.98 msec and 782.73 msec, respectively.

The Discriminability main effect reached significance because response times decreased as stimulus difference increased, but at a decreasing rate. The mean response times at Discriminability levels 1 to 5 were 907.54 msec, 836.44 msec, 762.78 msec, 747.54 msec and 709.95 msec, respectively.

The Probability x Response interaction, shown in table 10, arose because, in both types of block, times were shorter for responses to the more probable stimulus. Specifically, in Probability A, right responses were faster than left responses and, in Probability B, left responses were faster than right responses.

	PROBABILITY	
RESPONSE	A	B
Right	724.98	842.34
Left	838.74	757.25

TABLE 10. Mean times for correct right and left responses in blocks of Probability A and blocks of Probability B. The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

Figure 31 shows the interaction between Probability, Half and Response. From the figure it can be seen that times for responses to the more probable stimulus increased from Half 1 to Half 2 while times for responses to the less probable stimulus decreased from Half 1 to Half 2. In particular, in blocks of type A, times for right responses increased and times for left responses decreased from the first to the second half of the block. At the same time, in blocks of type B, times for right responses decreased and times

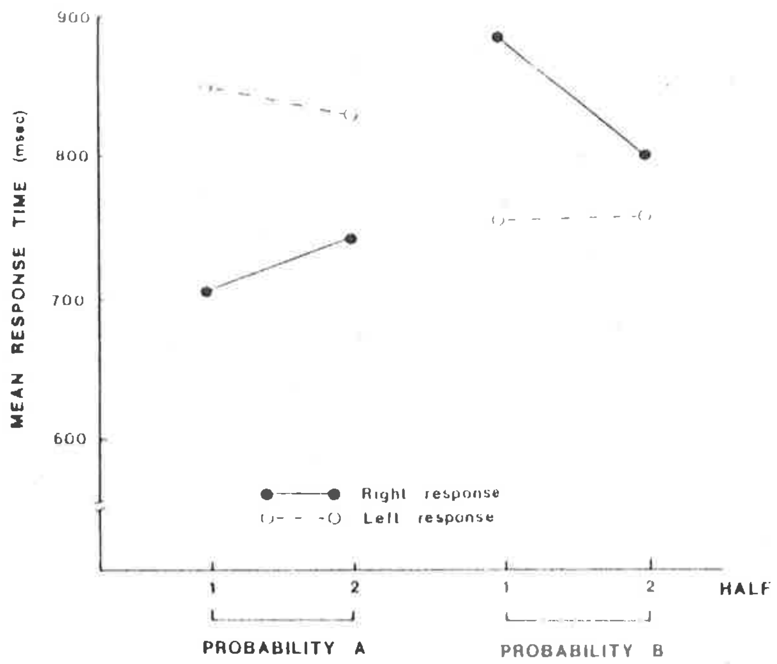


FIGURE 31. Mean times for correct right responses (solid line) and correct left responses (broken line) in Half 1 and Half 2 of blocks of Probability A (left pair of lines) and of Probability B (right pair of lines). The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

for left responses increased from the first to the second half of the block. However, the changes in times for left responses were not significant. In addition, there were significant changes in times across block boundaries. Times for right responses increased from Half 2 of Probability A to Half 1 of Probability B. Then, they decreased from Half 2 of Probability B to Half 1 of Probability A. Left responses showed the opposite pattern.

The interaction of Period x Probability x Half x Discriminability is shown in figure 32. Although this interaction was also significant for accuracy data, this result did not appear to indicate any clear trend in response time.

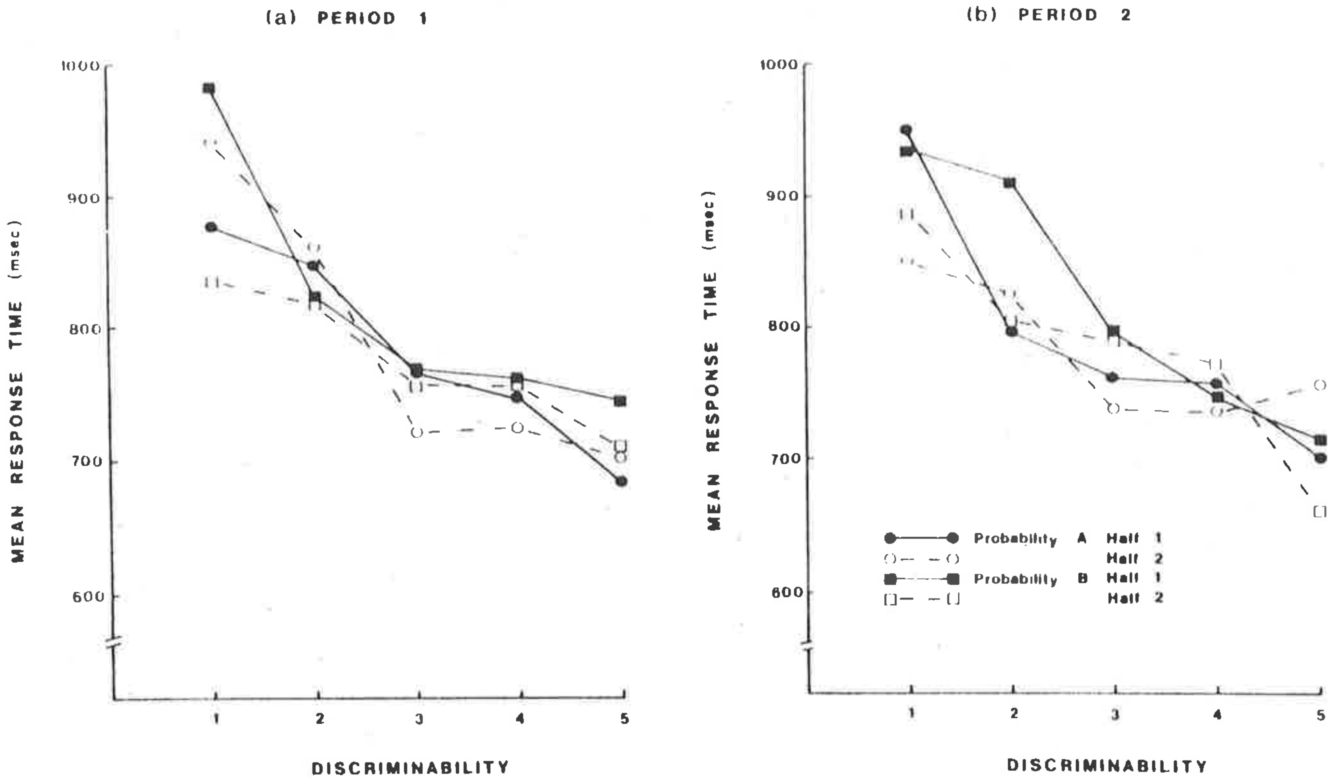


FIGURE 32. Mean times for correct responses at each level of discriminability in Half 1 (solid line) and Half 2 (broken line) of blocks of Probability A (circles) and of Probability B (squares) in (a) Period 1 and (b) Period 2. The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

Similarly, the interaction between Period, Half, Discriminability and Response, shown in figure 33, did not appear to result from any consistent variation in response times even though there was a corresponding result in accuracy data. Instead, it appeared to arise from apparently random fluctuations in response times between periods, especially at the easiest levels of Discriminability, levels 3, 4 and 5.

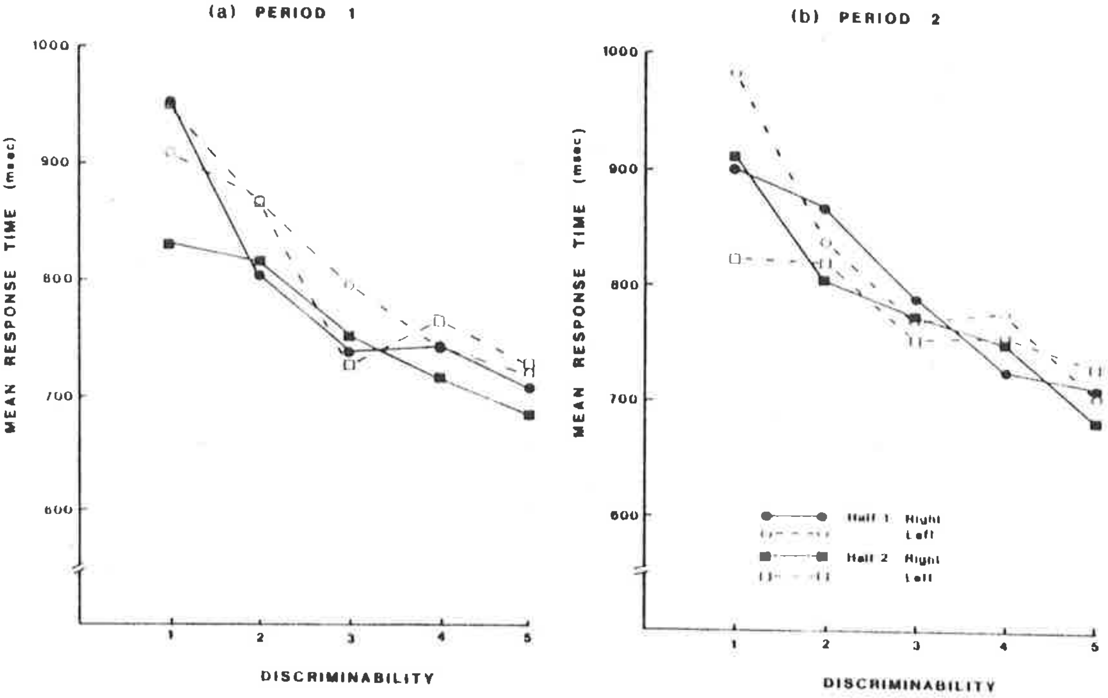


FIGURE 33. Mean times for correct right responses (solid line) and correct left responses (broken line) at each level of discriminability in Half 1 (circles) and Half 2 (squares) of blocks in (a) Period 1 and (b) Period 2. The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

The interaction of Probability x Half x Discriminability x Response x Order is shown in figure 34 where it is apparent that this interaction reached significance due to unsystematic changes in response times across levels of discriminability. However, from the figure it can be seen that, in blocks of

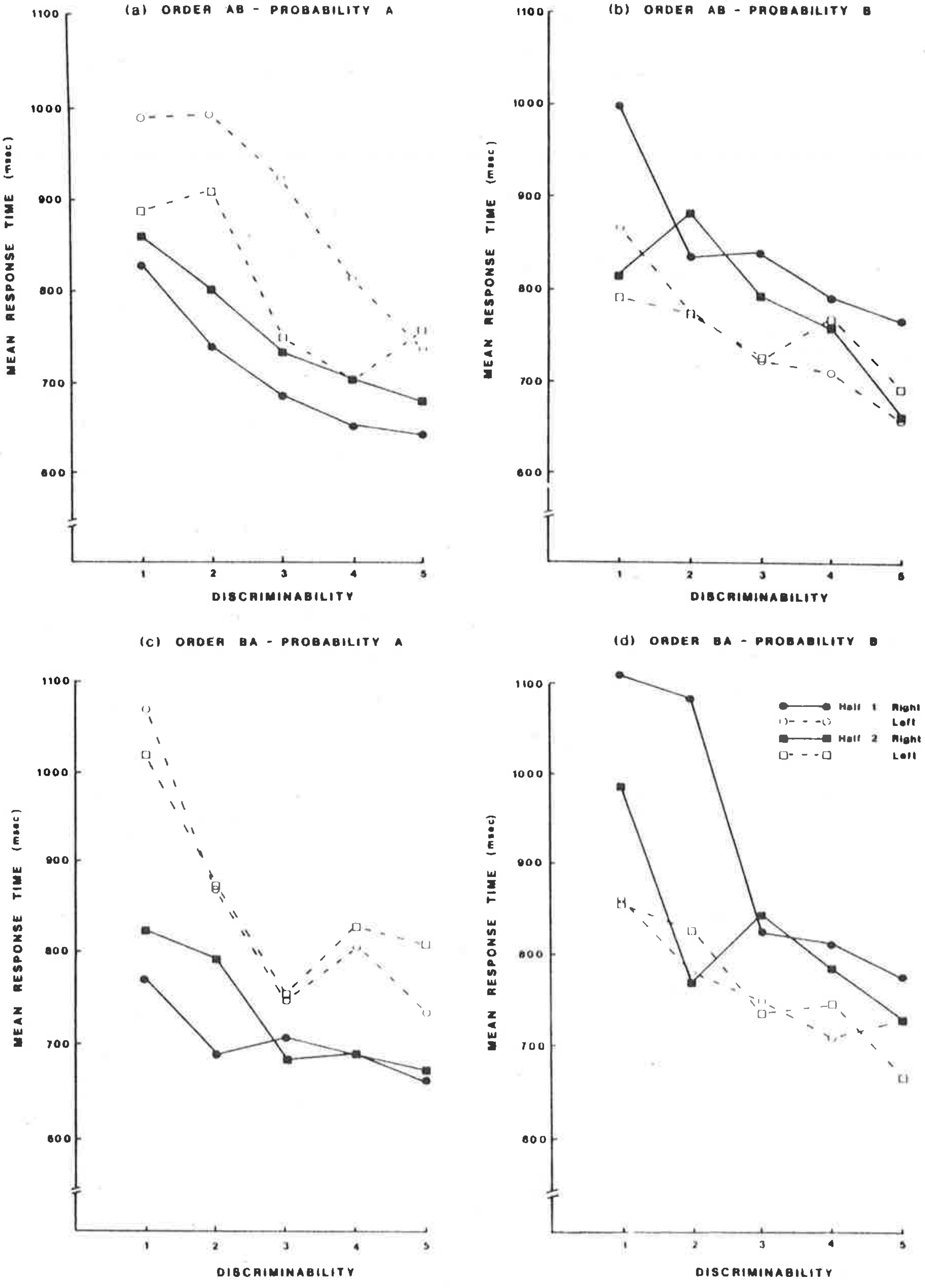


FIGURE 34. Mean times for correct right responses (solid line) and correct left responses (broken line) at each level of discriminability in Half 1 (circles) and Half 2 (squares) of blocks of (a) Probability A in Order AB, (b) Probability B in Order AB, (c) Probability A in Order BA and (d) Probability B in Order BA. The means were taken over the 10 observers in each order.

type A in both orders, left responses tended to be slower than right responses at all levels of discriminability. Similarly, in blocks of type B in both orders, right responses tended to be slower than left responses at all levels of discriminability. In addition, it appeared that responses in Order BA tended to be slower overall.

(c) Confidence

The analysis of confidence data showed significant main effects due to Half ($F(1,18) = 14.79$, MSerror = 0.40, $p < .01$) and Discriminability ($F(4,72) = 50.56$, MSerror = 0.99, $p < .001$). In addition there were significant interactions of Probability x Half ($F(1,18) = 5.07$, MSerror = 0.40, $p < .05$), Period x Discriminability ($F(4,72) = 2.82$, MSerror = 0.31, $p < .05$), Probability x Response ($F(1,18) = 9.09$, MSerror = 1.43, $p < .01$), Probability x Half x Order ($F(1,18) = 6.00$, MSerror = 0.40, $p < .05$), Period x Probability x Response ($F(1,18) = 4.52$, MSerror = 0.37, $p < .05$), Half x Response x Order ($F(1,18) = 4.70$, MSerror = 0.27, $p < .05$), Probability x Half x Response ($F(1,18) = 13.01$, MSerror = 0.42, $p < .01$), Period x Probability x Half x Discriminability ($F(4,72) = 3.95$, MSerror = 0.26, $p < .01$), Period x Discriminability x Response x Order ($F(4,72) = 2.52$, MSerror = 0.33, $p < .05$) and Period x Probability x Half x Discriminability x Response x Order ($F(4,72) = 2.86$, MSerror = 0.20, $p < .05$).

The Half main effect arose because confidence was lower in the first half of blocks than in the second half of blocks. Mean confidence in Half 1 and Half 2 was 2.79 and 2.92, respectively.

The Discriminability main effect reached significance because confidence increased with stimulus difference, but at a decreasing rate.

The mean confidence at Discriminability levels 1 to 5 was 2.30, 2.62, 2.92, 3.14 and 3.28, respectively.

The Probability x Half interaction is shown in table 11 where it can be seen that confidence was higher in Half 1 of blocks of type A than in Half 1 of blocks of type B. There was no difference between confidence in Half 2 of blocks of types A and B.

HALF	PROBABILITY A	 B
1	2.86	2.72
2	2.91	2.92

TABLE 11. Mean confidence in correct responses in Half 1 and Half 2 of blocks of Probability A and blocks of Probability B. The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

The Period x Discriminability interaction (see figure 35) reached significance because, at Discriminability level 1, confidence was higher in Period 1 than in Period 2. This trend was reversed at Discriminability level 2 where confidence was lower in Period 1 than in Period 2.

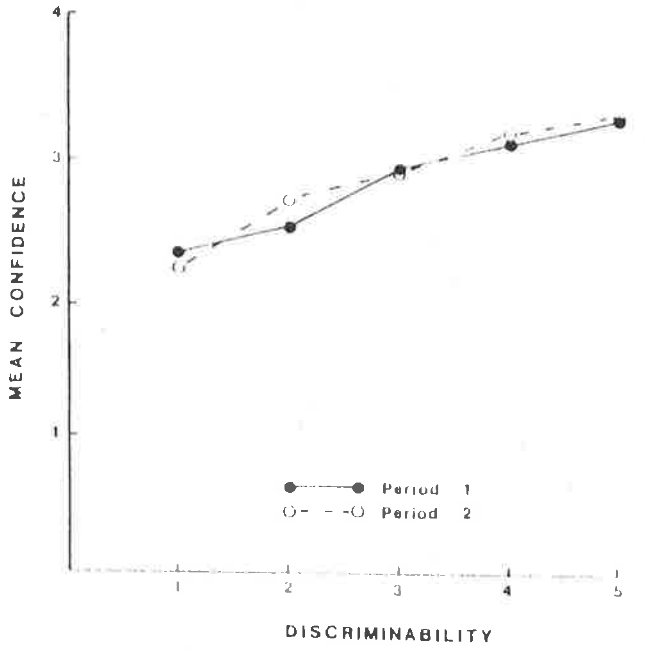


FIGURE 35. Mean confidence in correct responses at each level of discriminability in Period 1 (solid line) and Period 2 (broken line). The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

The Probability x Response interaction is shown in table 12 where it can be seen that, in both types of block, confidence was higher in responses to the more probable stimulus. In particular, in Probability A, confidence was higher in right responses than in left responses. Conversely, in Probability B, confidence was higher in left responses than in right responses.

Figure 36 shows the interaction between Probability, Half and Order. From the figure it is apparent that this interaction reached significance because there was a significant increase in confidence from the first to the second half of blocks of type B in Order BA. Confidence did not vary from the first to the second half of blocks in any other case.

RESPONSE	PROBABILITY A	 B
Right	3.01	2.76
Left	2.76	2.87

TABLE 12. Mean confidence in correct right and left responses in blocks of Probability A and of Probability B. The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

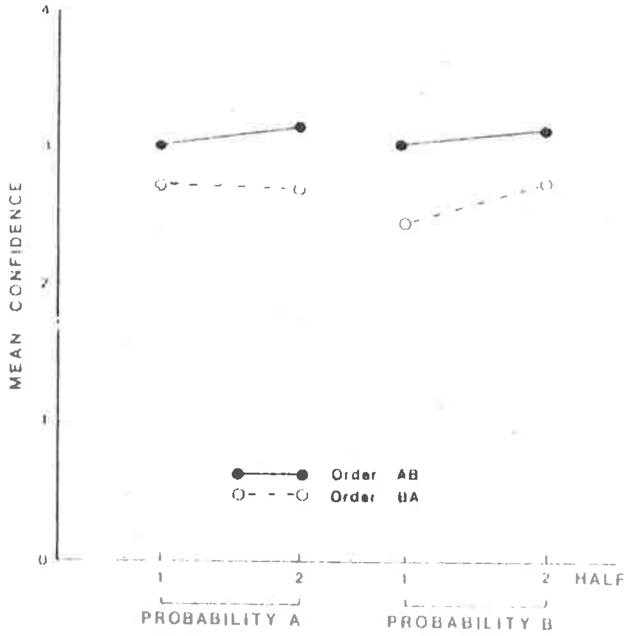


FIGURE 36. Mean confidence in correct responses in Half 1 and Half 2 of blocks of Probability A (left pair of lines) and of Probability B (right pair of lines) in Order AB (solid line) and Order BA (broken line). The means were taken over the 10 observers in each order.

The Period x Probability x Response interaction, shown in table 13, arose because there was no difference in confidence in right and left responses between periods in blocks of type A. In contrast, in blocks of type B, confidence in right responses decreased from Period 1 to Period 2, while confidence in left responses increased from Period 1 to Period 2. At the same time, in blocks of type A in both periods, confidence was higher for right responses than for left responses. In blocks of type B, in Period 1 there was no difference in confidence for right and left responses and in Period 2, left responses were more confident than right responses.

	PROBABILITY A		PROBABILITY B	
RESPONSE	PERIOD 1	PERIOD 2	PERIOD 1	PERIOD 2
Right	3.01	3.00	2.81	2.71
Left	2.75	2.76	2.78	2.96

TABLE 13. Mean confidence in correct right and left responses in blocks of Probability A and of Probability B in Period 1 and Period 2. The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

In figure 37 it can be seen that the Half x Response x Order interaction arose because, in Order AB, there was a greater increase in confidence in left responses than in confidence in right responses from Half 1 to Half 2. Conversely, in Order BA, confidence in right responses showed a greater increase from Half 1 to Half 2 than did confidence in left responses. In addition, confidence was higher overall in Order AB than in Order BA.

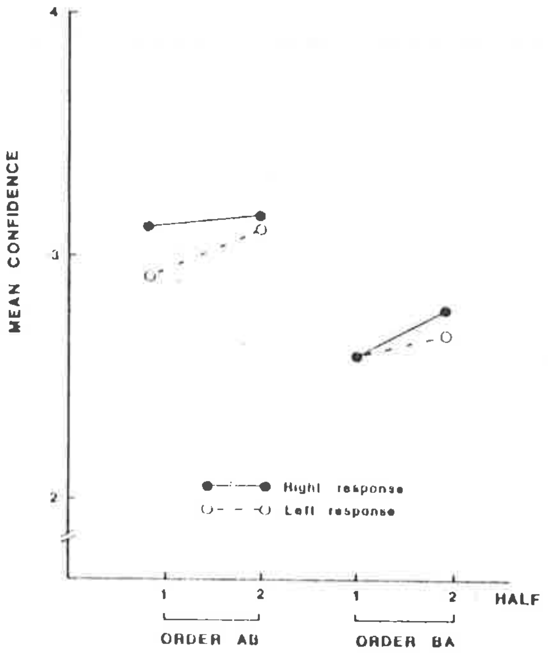


FIGURE 37. Mean confidence in correct right responses (solid line) and correct left responses (broken line) in Half 1 and Half 2 of blocks in Order AB (left pair of lines) and Order BA (right pair of lines). The means were taken over the 10 observers in each order.

Figure 38 shows the interaction between Probability, Half and Response. From the figure it can be seen that confidence increased from Half 1 to Half 2 for all except right responses in Probability A. In this case, confidence decreased from Half 1 to Half 2. Confidence in right responses decreased from Half 2 of Probability A to Half 1 of Probability B and increased from Half 2 of Probability B to Half 1 of Probability A. Confidence in left responses showed the opposite changes. None of these differences in confidence in either right or left responses was significant.

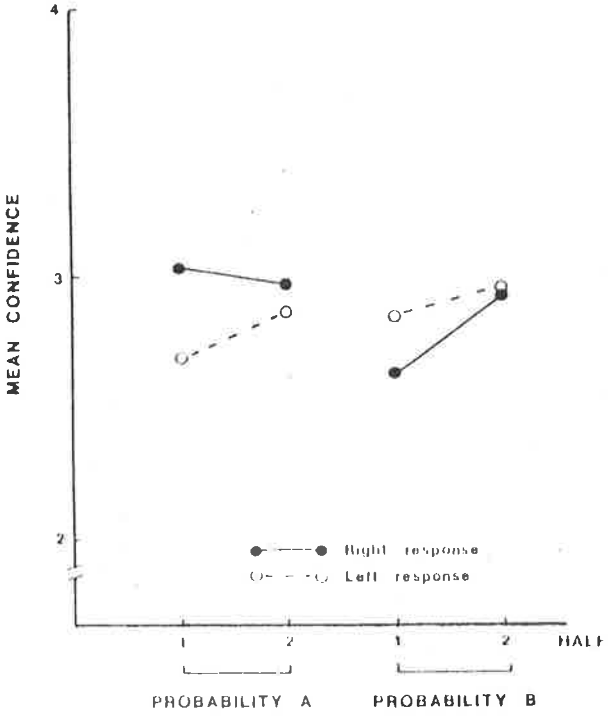


FIGURE 38. Mean confidence in correct right responses (solid line) and correct left responses (broken line) in Half 1 and Half 2 of blocks of Probability A (left pair of lines) and Probability B (right pair of lines). The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

From figure 39 it appears that the interaction between Period, Probability, Half and Discriminability was not indicative of any systematic changes in confidence. Instead, this interaction appeared to reach significance as a result of inconsistent fluctuations in confidence between Discriminability levels. Similarly, the interaction between Period, Discriminability, Response and Order (see figure 40) was not due to any systematic variations in confidence in right and left responses between periods in the different order conditions. However, confidence did appear to be lower overall in Order BA.

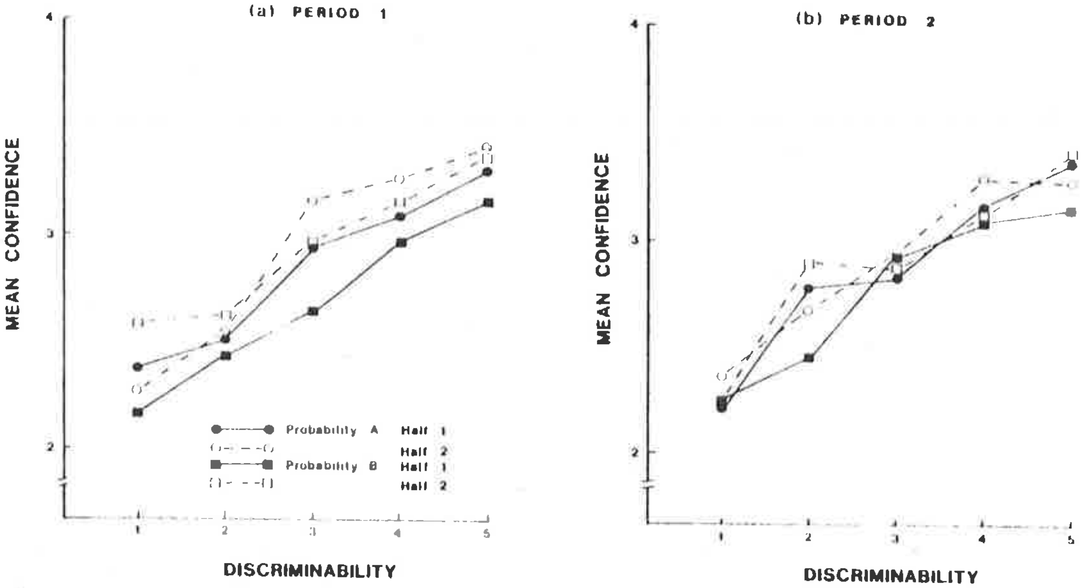


FIGURE 39. Mean confidence in correct responses at each level of discriminability in Half 1 (solid line) and Half 2 (broken line) of blocks of Probability A (circles) and of Probability B (squares) in (a) Period 1 and (b) Period 2. The means were taken over 20 observers, the 10 observers in Order AB and the 10 observers in Order BA.

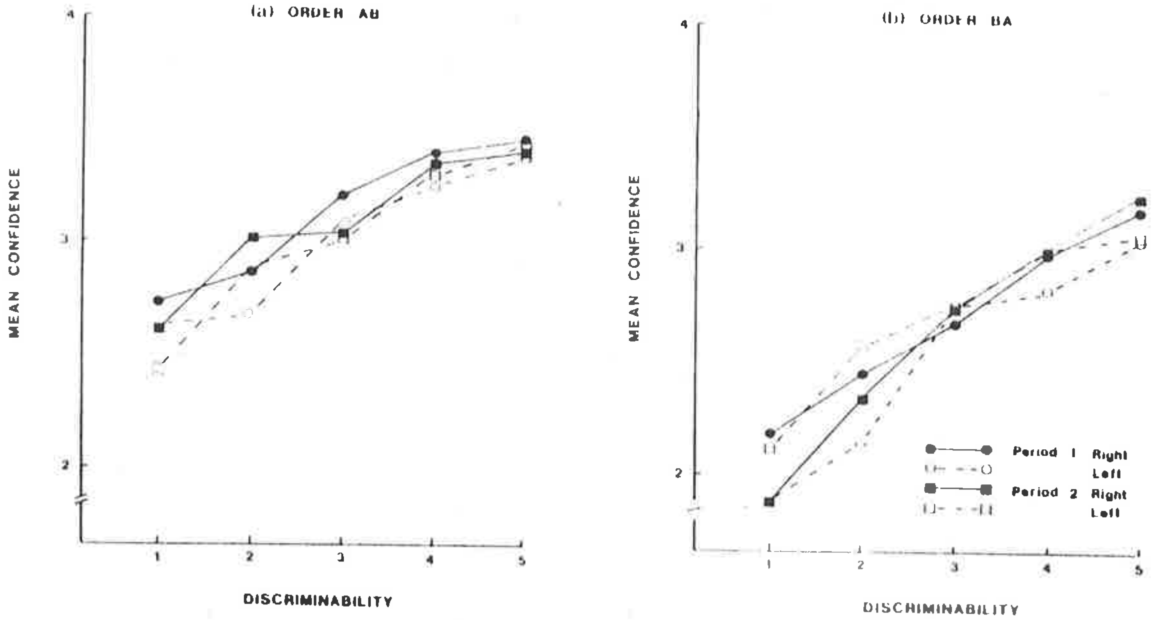


FIGURE 40. Mean confidence in correct right responses (solid line) and correct left responses (broken line) at each level of discriminability in Period 1 (circles) and Period 2 (squares) of (a) Order AB and (b) Order BA. The means were taken over the 10 observers in each order.

The 6-way interaction between Period, Probability, Half, Discriminability, Response and Order is shown in table 14. However, since this level of interaction is outside the usual range of predictions of any model, no attempt was made to interpret it.

		PERIOD 1				PERIOD 2			
PROBABILITY		A		B		A		B	
HALF		1	2	1	2	1	2	1	2
ORDER AB									
RESPONSE									
RIGHT	1	2.93	2.63	2.50	2.85	2.66	2.52	2.70	2.55
	2	3.01	2.88	2.75	2.85	3.06	3.06	2.90	3.10
	3	3.31	3.37	2.95	3.20	3.21	3.08	2.95	2.95
DISCRIMINABILITY	4	3.51	3.62	3.20	3.30	3.58	3.50	3.05	3.30
	5	3.62	3.59	3.35	3.35	3.53	3.62	3.10	3.35
LEFT	1	2.55	2.70	2.51	2.77	2.00	2.55	2.52	2.62
	2	2.25	2.75	2.77	2.90	2.90	2.75	2.87	2.99
	3	2.90	3.25	3.04	3.19	2.45	3.10	3.36	3.18
	4	3.10	3.35	3.27	3.32	2.90	3.45	3.51	3.35
	5	3.30	3.35	3.42	3.50	3.35	3.30	3.44	3.60
ORDER BA									
RIGHT	1	2.29	2.08	1.80	2.60	2.30	2.11	1.50	1.45
	2	2.62	2.55	2.15	2.45	2.73	2.38	1.45	2.75
	3	2.75	2.80	2.35	2.75	2.99	2.92	2.65	2.45
DISCRIMINABILITY	4	3.02	3.04	2.55	3.00	3.22	3.06	2.85	2.90
	5	3.31	3.17	2.70	3.55	3.27	3.34	2.80	3.50
LEFT	1	1.80	1.60	1.85	2.12	1.85	2.10	2.25	2.26
	2	2.15	2.05	2.13	2.24	2.50	2.50	2.62	2.67
	3	2.90	3.10	2.25	2.79	2.65	2.75	2.84	2.80
	4	2.70	2.95	2.75	2.87	2.95	3.10	2.94	3.04
	5	2.85	3.35	3.00	2.99	3.20	2.80	3.16	3.07

TABLE 14. Mean confidence in correct right and left responses at each level of discriminability in each half of blocks of Probability A and Probability B in both Period 1 and Period 2. Means are shown for Order AB and for Order BA. The means were taken over the 10 observers in each order.

(d) Confidence and response time

For each observer in each group, confidence ratings for both correct and incorrect responses were plotted against the corresponding values of response time, and straight lines were fitted. This was done for right and left responses separately and for each level of discriminability in Probability A and in Probability B. There were 400 plots. It was found that 289 of the slopes of the best fitting straight lines were negative, showing that overall, there was an inverse relationship between confidence and response time ($p < .001$, binomial test). The inverse relationship was still in evidence when right and left responses at each level of discriminability were considered for each type of block. Table 15 shows the number of negative slopes of the 100 best fitting straight lines in each case. Here it can be seen that, in all cases, the majority of the best fitting lines had a negative slope.

PROBABILITY	NEGATIVE SLOPES RIGHT RESPONSES	P	NEGATIVE SLOPES LEFT RESPONSES	P
A	79	<.001	76	<.001
B	70	<.001	76	<.001

TABLE 15. The number of negative slopes obtained when, for each of the 10 observers in each order condition, confidence ratings were plotted against the corresponding response time values and straight lines were fitted. Separate lines were fitted for right responses at each level of discriminability in blocks of Probability A and of Probability B and for left responses at each level of discriminability in blocks of Probability A and of Probability B. Therefore, the maximum possible number of negative slopes for each response in each probability condition was 100. The associated binomial probability values are also shown.

4. Discussion

As in Experiment 3, the effects of the variations in discriminability which were observed in this experiment are consistent with the predictions of any of the mechanisms for criterion regulation. On the other hand, the manipulation of bias produced some results which cannot be explained in terms of any of the processes. However, it appears that many of these results were not due to the experimental manipulation. Instead, these findings appear to have arisen from the division of the data from the experimental session into blocks of 50 trials. Random changes within each block of trials of the experimental session appear to have produced significant effects which do not reflect the experimental manipulation and which do not correspond to simple predictions of any of the decision mechanisms. Therefore, only main effects and 2-way and 3-way interactions will be considered.

(a) Accuracy

The interaction between Probability, Half and Response indicated that there was a change in the relative accuracy of right and left responses both across block boundaries, where a priori probability changed, and within blocks of trials in which a priori probability was constant. Since the particular responses varied according to the probability of the associated stimulus, these changes appear to arise from the manipulation of a priori probability, but it was not apparent whether the change in performance in response to the step change in probability was indicated by the measure in the first or the second half of a block.

If the measure in the first half of a block indicated the effects of the change in probability then an increase in the a priori probability of a stimulus produced an increase in accuracy, while a decrease in a priori probability produced a decrease in accuracy. This finding contrasts with the results of Experiment 3. However, it is consistent with Laming's (1968, 1969) findings and can be explained in terms of the ideal observer hypothesis applied to any of the models if it is assumed that the observer uses a measure of the cumulative probability of right responses and of left responses as an estimation of the actual stimulus probabilities. For example, when there is a step increase in the probability of a right (or signal) response, β will decrease so that more right responses will be made and accuracy in right responses will increase. The ideal observer hypothesis would not predict the subsequent decrease in accuracy as the block progresses. However, in Experiment 3, accuracy data suggested that observers made an assessment of the stimulus probabilities which was not as extreme as the actual probabilities. It appeared that, in order to compensate for the apparently inappropriately large number of responses of one kind which they were making when compared with their probability assessment, observers would adjust the criterion for this response upwards. This form of compensatory criterion adjustment would explain the change in accuracy from the first to the second half of blocks in the present experiment and is consistent with all of the mechanisms of criterion regulation.

These data are not consistent with the response stabilisation hypothesis according to which a step increase in the probability of a stimulus will mean that the local probability will exceed the cumulative probability for the corresponding response so that there will be an increase in the criterion for that response. As a result, fewer responses will be made to the more probable stimulus and accuracy will be lower. This is the opposite of the pattern shown by data in the first half of blocks. Similarly, the adaptation-level hypothesis predicts that, in those blocks in which for example, right stimuli are more probable, in terms of Signal Detection Theory, the cutoff will move so that fewer right responses will be made and accuracy will decrease. When these mechanisms are applied to the random walk or accumulator models the qualitative predictions will be the same as those for Signal Detection Theory.

On the other hand, these data are predicted by the target confidence mechanism. For example, in terms of the accumulator model, when there is a step increase in the probability of one stimulus, the criterion for the corresponding response will decrease relative to the criterion for the other response and, as a consequence, accuracy for responses to the more probable stimulus will increase. Since this control mechanism reacts quickly to any change in the stimulus sequence (for example, see Vickers, 1979, fig. 143(a)) it would be expected that the response to a change in probability would be evident in the first half of a block. In addition, simulations of the process show that, over time, the change in criteria in response to a step change in probability takes the form of a series of decreasing oscillations (see Vickers, 1979, fig. 143(a)). In other words, after criterion values diverge they tend to converge again. The decrease in the accuracy of responses to the more probable stimulus in the second half of a block would be predicted as part of such a reversal. It appears that the reversal is interrupted by the next step change in probability.

These data appear to contradict the findings in Experiment 3. However, from the trend shown by accuracy across blocks in Experiment 3 it appeared that observers were adjusting their criterion values in accordance with subjective assessments of stimulus probability, which were becoming more accurate as the session progressed. Indeed, it appeared that, if observers had performed several extra blocks, the pattern would have resembled that found in the present experiment. In addition, in both cases, the results can be accommodated by the models in terms of variations in criteria. Therefore, it seems likely that this conflict indicates a difference in the rate of adjustment of criteria in the different experimental situations rather than a different response to the change in probability. It appears that, in Experiment 3, data from the entire experimental session comprised a single adaptation to the stimulus probabilities, while in Experiment 4 observers adapted to different values of a priori probability in each block.

Alternatively, if it is assumed that the measure in the second half of the block represented the change in response to the variation in a priori probability, these data showed that accuracy decreased as a priori probability increased. This is the same pattern as was observed in Experiment 3, but it is contrary to the findings of Laming (1968, 1969) and conflicts with the ideal observer hypothesis and the target confidence mechanism. However, as described above, this pattern is predicted by the response stabilisation and adaptation-level hypotheses.

The data are most easily explained if it is assumed that the measure in the first half of the block was the response to the change in probability. The change in accuracy during the block may indicate that the observer's assessment of the probabilities was not appropriate and that he was making compensatory adjustments in criteria to try to reduce the discrepancy between his expectations and his responding. On the other hand, it may indicate a reversal in the regulatory control of a negative feedback system as in the accumulator model due to an initial overshoot.

Alternatively, if it is assumed that the measure in the second half was the response to the change in probability, then the measure in the first half is best interpreted as a continuation of the trend in the previous block. However, this does not appear to be the case in these data, where the change in accuracy across block boundaries tended to be steeper than the change within blocks.

(b) Response time

The main effect of Half and the interactions of Probability x Response and Probability x Half x Response all appear to arise from the same variation in response time and will be discussed together in terms of the 3-way interaction. Times for responses to the more probable stimulus increased and times for responses to the less probable stimulus decreased from the first to the second half of blocks. Once more it is not clear whether the response to the change in probability was shown by the measure in the first or the second half of the block.

If the measure in the first half is interpreted as the response to a priori probability, response times decreased when probability increased. This corresponds to Laming's (1968, 1969) findings and can be explained in terms of the ideal observer hypothesis in Signal Detection Theory by the choice of an appropriate latency function. These data are also consistent with the ideal observer hypothesis applied to the random walk or the accumulator model. In both cases, according to the ideal observer hypothesis, the criterion amount of information required for responses to the more probable stimulus will be low so that these responses will tend to be fast. As for accuracy data, the subsequent change in times during the block could indicate a compensatory adjustment in criteria by observers whose assessment of stimulus probabilities did not match their responding.

These data are not consistent with the response stabilisation or adaptation-level hypotheses which, for example, in terms of Signal Detection Theory, predict that the criterion for the response to the more probable stimulus will be high, so that times for this response will be long. Predictions of these mechanisms applied to the random walk and accumulator models are qualitatively identical to those for Signal Detection Theory.

However, these data are predicted by the target confidence mechanism applied to the random walk or accumulator models, since the low criterion for responses to the more probable stimulus will give rise to fast responses. Once more, this process would predict that reversals in the adjustment of criteria during the block would produce corresponding reversals in response time. Therefore, the model predicts the increase in times for the more probable response in the second half of blocks.

If the measure in the second half is interpreted as the response to a priori probability, response times increased with probability. This contrasts with Laming's (1968, 1969) results but it can be explained by the response stabilisation hypothesis according to which the criterion for responses to the more probable stimulus will be adjusted upwards because the local probability of this response exceeds its cumulative probability. Alternatively, the adaptation-level hypothesis predicts that the indifference point will be adjusted to correspond to the mean of all observations so that the discriminability of the more probable stimulus is reduced. In both cases, the adjustment will result in longer times for responses to the more probable stimulus.

As in accuracy data, it appears more likely that the measure in the first half of blocks represented the change due to a priori probability because the changes in response times across block boundaries tended to be steeper than the changes within blocks.

(c) Confidence

As with response time data, the main effect of Half and the interactions of Probability x Half, Probability x Response and Probability x Half x Response appear to result from the same effect, so they will all be discussed together in terms of the Probability x Half x Response interaction. In general, confidence in both right and left responses increased from the first to the second half of blocks while there were no consistent changes in confidence between blocks. When this result is considered in conjunction with response time data it is apparent that confidence was not always varying inversely with response time. In particular, for left responses in blocks of type B there was a direct relationship between confidence and response time. Therefore, these data are not consistent with Audley's (1960) hypothesis, so they cannot be explained by any of the mechanisms for criterion regulation applied to Signal Detection Theory or the random walk model. In addition, when this finding is considered in terms of the balance of evidence hypothesis of the accumulator model, some problems arise. In order to be able to account for accuracy and response time data in terms of the target confidence mechanism it was necessary to assume that the response to the change in probability was represented by measures in the first half of blocks. In this way, high accuracy and short times for responses to the more probable stimulus would indicate a low criterion for that response. Therefore, the balance of evidence hypothesis would predict that confidence in this response will be low. This is the opposite of the data in the first half of blocks. Conversely, the low accuracy and long times for responses to the less probable stimulus should correspond to a high criterion and therefore confidence should be high. This prediction did not receive any support from the confidence data. Similarly, in contrast to these data the adaptation-level hypothesis predicts that confidence in responses to the more probable stimulus will be lower than confidence in responses to the less probable stimulus.

When the response stabilisation hypothesis is applied to the accumulator model, the criterion for responses to the more probable stimulus will be high so that confidence will be high. This is the trend found in the first half of blocks. Conversely, this mechanism predicts that confidence for responses to the less probable stimulus will be low, as seen in the first half of blocks. However, as seen above, this mechanism could not account for accuracy or response time results.

If the measure in the second half of blocks indicated the response to a change in probability, the ideal observer hypothesis, the target confidence mechanism or the adaptation-level hypothesis can explain the data in blocks of type A where confidence in responses to the more probable stimulus is relatively low. However, none of the mechanisms can account for the findings in blocks of type B where confidence in responses to both stimuli increases from Half 1 to Half 2. Therefore, it appears that when all aspects of the confidence data are considered together they are not consistent with any one of the mechanisms.

The Period x Discriminability interaction appears to result from a difference in confidence in right and left responses at Discriminability level 2. In this way, it is not a specific prediction of any of the models. However, it does not conflict with the basic form of any of the models.

Similarly, the higher order interactions obtained with Order (Probability x Half x Order, Half x Response x Order) are not inconsistent with any of the models but they do not correspond to simple predictions of the models.

At the same time, although it is not apparent how the Period x Probability x Response interaction could be derived from the models, this interaction does not conflict with the models as they stand.

(d) Confidence and response time

When the relationship between confidence and response time was considered within bias conditions, confidence varied inversely with time.

However, as in Experiment 3 it appears that this was not necessarily the form of the relationship between confidence and response time between bias conditions.

5. Summary and Conclusions

For each dependent variable, the variations in discriminability produced systematic effects which were consistent with all of the models. However, the effects of *a priori* probability were much less well defined, especially in confidence data.

A major problem arose when considering the effects of the manipulation of bias. In each dependent variable, the response to the step change in probability varied from the first to the second half of blocks and there was little to indicate which measure represented the initial response to the change in probability. Furthermore, the accuracy and response time measures from the first and the second half of blocks could not be accommodated equally well by the models. None of the models can explain confidence measures in either the first or the second half of blocks in terms of the effects of bias.

It appeared likely that the measure in the first half of blocks indicated the change due to the variation in probability because the changes in the dependent variables across the block boundaries were more marked than the changes within the blocks. In this case, accuracy and response time data are consistent with the ideal observer hypothesis, the target confidence mechanism and the adaptation-level hypothesis. However, the measures in the first half of blocks in the present experiment showed a trend opposite to that found when accuracy measures were compared across conditions in Experiment 3. It is possible that the trend shown in Experiment 3 merely represented the effects of a between-subjects design with individual differences within the different groups of observers giving rise to this result. However, it appears more likely that this finding stems from the slow rate at which observers adapted to the stimulus probabilities in Experiment 3 in comparison to the rapid adaptations which appeared to occur here.

Alternatively, it could be the second measure in each block which indicates the initial response to the change in probability. If this is the case, accuracy and response time data are consistent with the response stabilisation hypothesis. In addition, this would imply that the measure in the first half of the block was a continuation of the trend from the previous block. This seems unlikely in the light of the finding that the change in performance across block boundaries tended to be steeper than the change within blocks.

In order to clarify this situation it would be necessary to examine in more detail the change in performance in response to a step change in probability. In particular, it appears that the change which occurs between measures taken before and after a step change in probability may indicate the speed with which any adaptation occurs. In addition, it appears that the inclusion of different levels of stimulus difference increased the variability of the data and may have obscured the pattern in confidence ratings. As in Experiment 3, the lack of clarity in confidence measures gave rise to findings which suggest that confidence varied directly with response time between bias conditions while there was strong evidence that there was an inverse relationship between confidence and response time within bias conditions. This apparent ambiguity might be resolved if confidence measures were taken at only one level of discriminability.

CHAPTER 5

A. INTRODUCTION

In Chapter 2 it was noted that each mechanism for criterion regulation in two-category tasks can be applied to each decision model, except in the case of the target confidence mechanism and Signal Detection Theory. As a preliminary examination of these processes in a situation in which a priori probabilities were varied, Experiments 3 and 4 were performed to investigate the pattern of empirical evidence. However, it was not possible to distinguish adequately between the processes on the basis of these data alone.

The way in which each mechanism for criterion regulation can be applied to the three decision models was described in Chapter 2. The detailed predictions of each model for accuracy, response time and confidence, when there are step changes in a priori probability will now be examined. In particular, predictions are considered for the situation in which the a priori probability of right stimuli, p(R), alternates between 0.8 and 0.2 in successive blocks of trials, as in Experiment 4. Since a step increase in the probability of a right stimulus (or signal) necessarily corresponds to a step decrease in the probability of a left stimulus, predictions for responses to left stimuli vary in the complementary way to those for right stimuli, and detailed predictions will be presented in terms of responses to right stimuli only. Similarly, for all response measures, the change which results from a step decrease in probability is the converse of that from a step increase, so only the case of a step increase in probability will be considered in detail. In the specification of the predictions of each of the mechanisms it will be assumed that, as in Experiment 4, 300 practice trials in which p(R) = 0.5 and 100 practice trials in which p(R) = 0.2 were performed prior to the first experimental block. Following the design of Experiment 4, it will be assumed that p(R) takes the values of 0.8, 0.2, 0.8 and 0.2 in the four subsequent blocks each of 100 trials. Graphical illustrations of the predictions show these four blocks of trials which are numbered 1, 2, 3 and 4, respectively. It will be assumed that d' = 2.0, corresponding to the calculated value from data for discriminability level 2 in Experiment 4. To avoid the complication of negative intensities, the means of the normal distributions of sensory effect for left and right stimuli will be assumed to take the values 10 and 12, respectively, with $\sigma = 1.0$ for both distributions.

1. The ideal observer hypothesis

(a) Signal Detection Theory

From equation (4) it can be seen that, according to the ideal observer hypothesis, the value of the cutoff varies with changes in a priori probability and in costs and payoffs. If costs and payoffs are constant, it can be assumed that step changes in stimulus probabilities will produce changes in the position of the cutoff at a rate determined by the nature of the subjective probability estimates. These may be based on either (i) recent stimulus probability (calculated, for example, over the last x = 10 trials), or (ii) cumulatively determined stimulus probability (calculated over the entire experimental session, including practice trials).

(i) recent probability. If the ideal observer hypothesis is based on recent stimulus probability calculated over the last x trials, a step increase in p(R) from 0.2 to 0.8 will produce a step decrease in the expected value of β from p(n)/p(s) = (0.8/0.2) = 4.00 to (0.2/0.8) = 0.25 within x trials. This corresponds to a movement of the cutoff from x_{c_1} to x_{c_2} in figure 41. It is apparent that the rate at which this change in β takes place depends on the size of x, with smaller values of x giving rise to more rapid adaptation. For example,

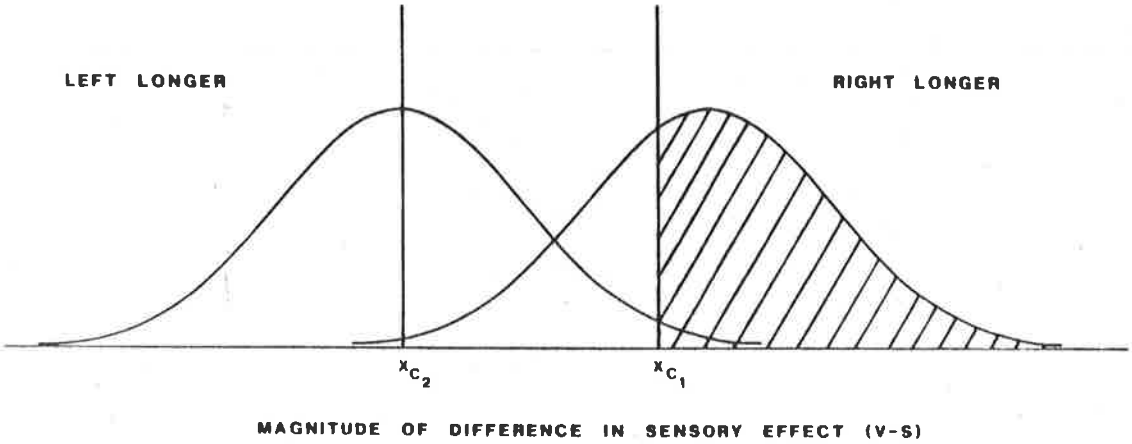


FIGURE 41. Theoretical probability density distributions of subjective values of stimulus difference (V-S) appropriate to Signal Detection Theory for two-category tasks. The right-hand distribution represents the case where the right line is longer and the left-hand distribution represents the case where the left line is longer. Two possible positions of the cutoff are shown. At the position x_{c_1}, responses 'left longer' are preferred, while at position x_{c_2}, responses 'right longer' are preferred. Therefore, the case when p(R)=0.2 and 'left longer' responses are preferred would be represented by cutoff x_{c_1}. An increase in p(R) from 0.2 to 0.8 with a concomitant increase in bias towards 'right longer' responses would correspond to a movement of the cutoff from x_{c_1} to x_{c_2}.

if x = 4, the adaptation of β described above will be completed by the end of the first four trials after a step increase in p(R). Alternatively, if x = 10, the adaptation of β will be completed by the end of the first 10 trials. Once the adaptation has taken place, the value of β (or the position of the cutoff) will remain constant until there is another change in p(R). The variation in the expected value of β based on recent stimulus probabilities calculated when x = 10 is shown in figure 42(a).

From figure 41 it is apparent that, when p(R) increases from 0.2 to 0.8, and the expected value of β decreases from 4.00 to 0.25, observations will be more likely to fall on the side of the cutoff corresponding to right responses. As a consequence, responses to right stimuli will be more accurate. For example, when p(R) = 0.2 so that the expected value of β = 4.00, from Freeman's (1964) tables it can be seen that the probability of correct responses to right stimuli is constant at 0.62. Conversely, when p(R) = 0.8 so that the expected value of β = 0.25, the probability of correct responses to right stimuli is constant at 0.95. Figure 42(b) illustrates the expected variations in the accuracy of responses to right stimuli which correspond to the changes in β shown in figure 42(a).

In Chapter 2 it was noted that, in terms of Signal Detection Theory, response time may be assumed to be an inverse function of the distance between an observation and the cutoff. The average position of observations which favour right responses can be estimated by the mean of the truncated normal distribution corresponding to correct right responses (the shaded area in figure 41). Therefore, the distance between this mean value and the cutoff will correspond to the average distance of observations favouring right responses from the cutoff. Specific estimates of this distance may be calculated for the experimental situation described above (Pearson & Hartley, 1976). In this case, when p(R) = 0.2, the average distance in standard deviation units, σ, between observations favouring right responses and the cutoff is 0.93. Conversely, when p(R) = 0.8, the average distance between observations favouring right responses and the cutoff is 1.80σ. Since response time varies inversely with this distance, the predicted response times will

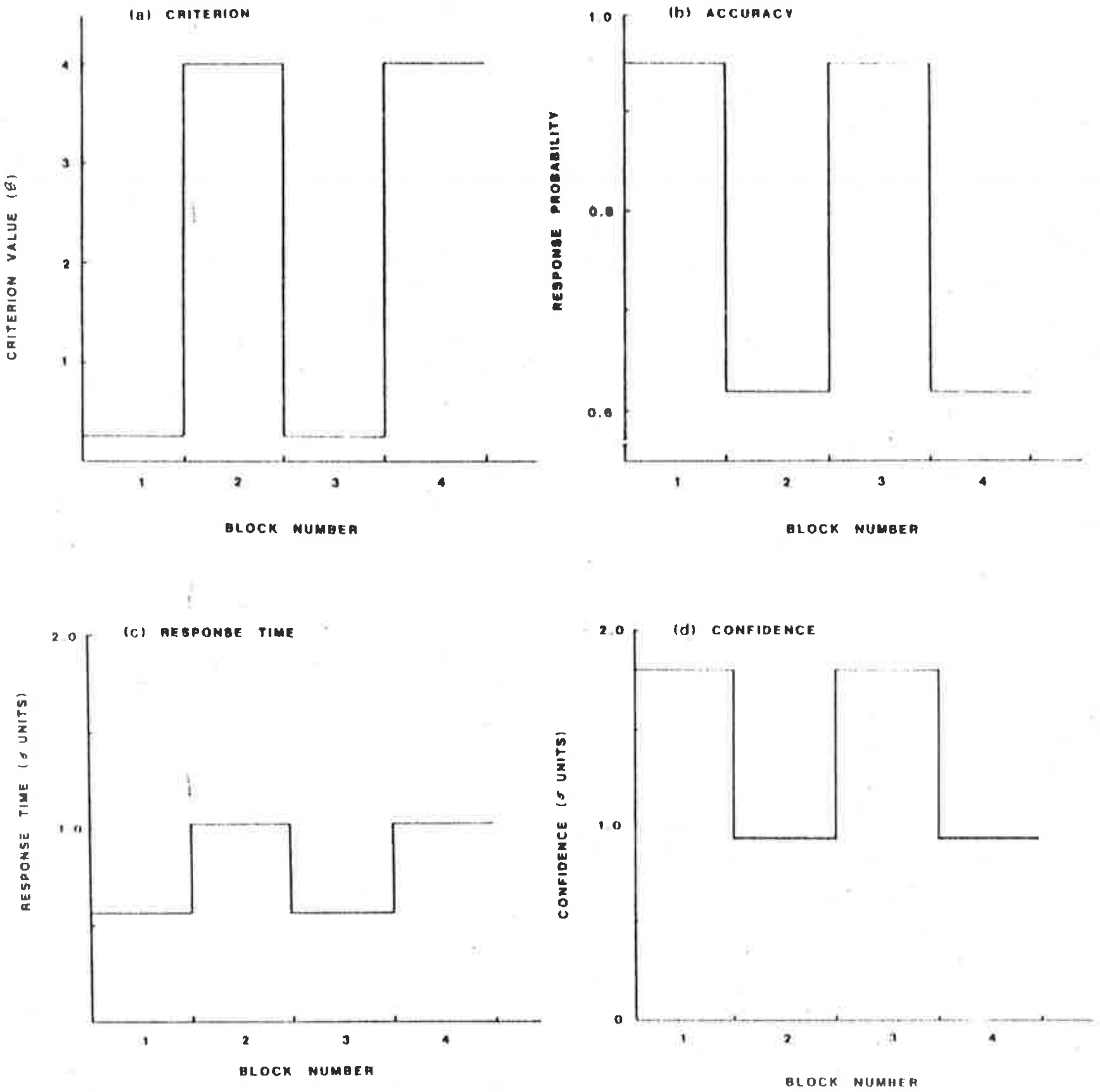


FIGURE 42. Predictions of the ideal observer hypothesis applied to Signal Detection Theory when the probability of a right stimulus, p(R), takes the values 0.8, 0.2, 0.8 and 0.2 in blocks of 100 trials which are numbered 1, 2, 3 and 4, respectively. It is assumed that the ideal observer hypothesis is based on recent stimulus probability, calculated over the last x = 10 trials. In addition, it is assumed that 300 practice trials in which p(R)=0.5 and 100 practice trials in which p(R)=0.2 preceded the trials which are plotted and that the distributions of sensory effect of left and right stimuli are normal with means of 10 and 12, respectively, with σ=1.0 for both distributions and d'=2.0. The figures show the expected variation in (a) the criterion, β, (b) response probability, (c) response time and (d) confidence.

vary in the same way as the reciprocal of these values as shown in figure 42(c). Times for right responses will be longer when p(R) = 0.2 than when p(R) = 0.8.

In Chapter 2 it was also noted that confidence may be assumed to vary as a direct function of the distance between an observation and the cutoff. As seen above, when p(R) = 0.2, the average distance between observations favouring right responses and the cutoff is 0.93σ, while when p(R) = 0.8, this average distance is 1.80σ. Therefore, as shown in figure 42(d), the predicted confidence in right responses will be lower when p(R) = 0.2 than when p(R) = 0.8.

(ii) cumulatively determined probability. Alternatively, β may depend on the stimulus probabilities cumulatively determined over the entire experimental session. In this case, the expected changes in β are in the same direction as those described above, but are more gradual and continue throughout the subsequent block. For example, figure 43(a) shows the variations in the expected value of β, based on cumulative stimulus probability, which occur in response to step changes in p(R) in the experimental situation described above. There is a ramp decrease in β following a step increase in p(R). From the figure it is apparent that the size of the change in the expected value of β decreases with each successive step change in p(R). This occurs because the cumulatively determined probability varies by smaller amounts as the number of trials over which it is calculated increases.

These changes in β will give rise to changes in all three response measures and, since the size of the variation in the expected value of β decreases with successive step changes in p(R), the magnitude of the expected changes in the response measures will decrease accordingly. For example, figure 43(b) shows the variations in the probability of correct responses to

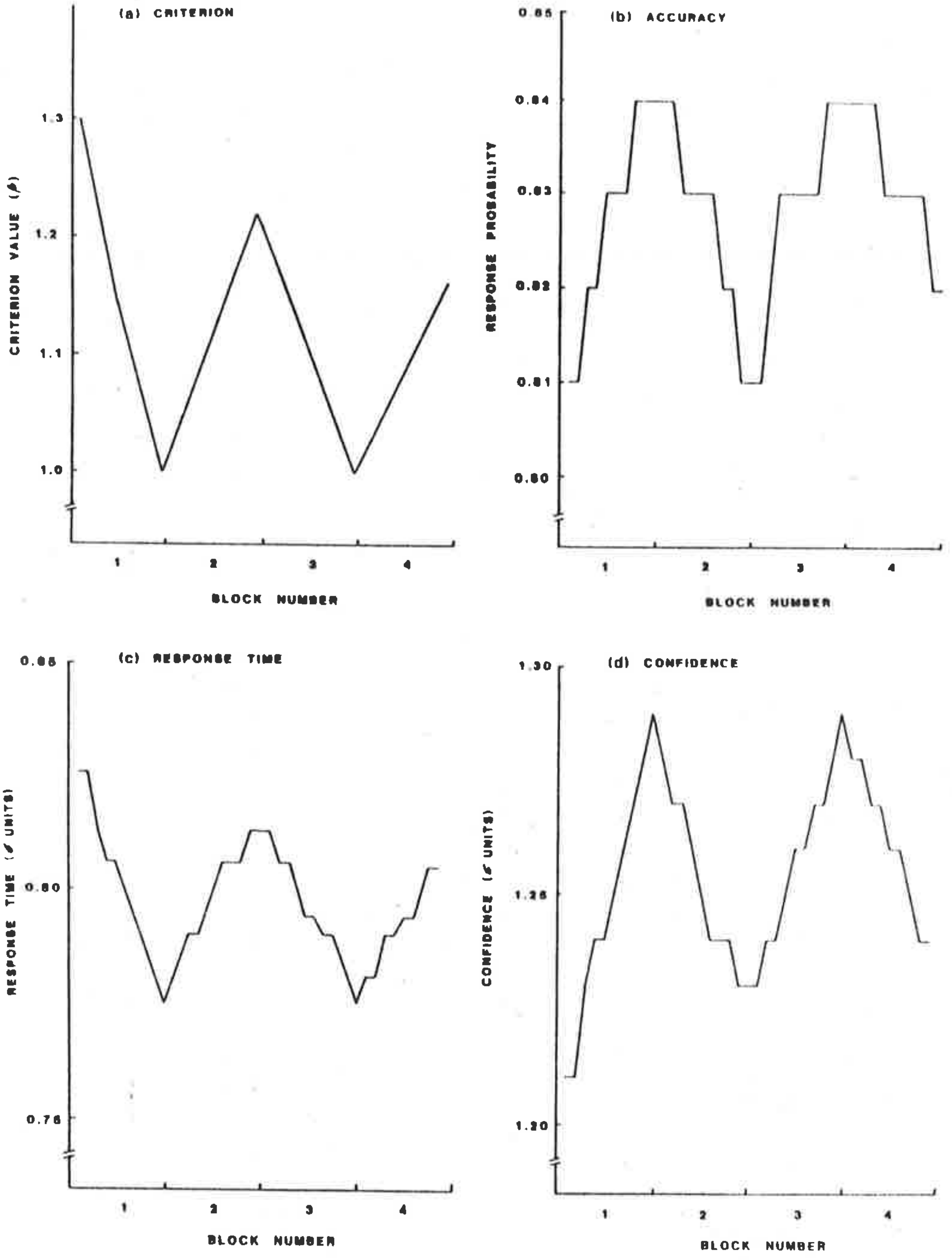


FIGURE 43. Predictions of the ideal observer hypothesis applied to Signal Detection Theory when the probability of right stimuli, p(R), takes the values 0.8, 0.2, 0.8 and 0.2 in blocks of 100 trials which are numbered 1, 2, 3 and 4, respectively. It is assumed that the ideal observer hypothesis is based on the stimulus probabilities cumulatively determined over the entire experimental session, and that 300 practice trials in which $p(R)=0.5$ and 100 practice trials in which $p(R)=0.2$ preceded the trials which are plotted. In addition, it is assumed that the distributions of sensory effect of left and right stimuli are normal with means of 10 and 12, respectively, with $\sigma=1.0$ for both distributions and $d'=2.0$. Predicted values were calculated every 10 trials.

right stimuli which correspond to the expected values of β plotted in figure 43(a). A step increase in p(R) gives rise to a ramp increase in the probability of correct responses to right stimuli which begins soon after the step change and continues throughout the block.

Secondly, as noted above, response time may be assumed to vary as an inverse function of the distance between an observation and the cutoff. As described above, the average distance between observations which favour right responses and the cutoff can be estimated by subtracting the cutoff value from the mean of the truncated normal distribution for correct right responses. The expected changes in response time values can be seen from a plot of the reciprocal of these differences which is given in figure 43(c). The figure shows that times for correct right responses decrease soon after a step increase in p(R) and continue to decrease throughout the block.

Thirdly, since confidence may be assumed to vary as a direct function of the distance between the observation and the cutoff, it will vary as shown in figure 43(d).

(b) The random walk model

When the prior odds aspect of the ideal observer hypothesis is applied to a random walk type of process, it is assumed that changes in a priori probability will produce variations in the position of the starting point. For example, if subjective probability is determined on the basis of recent stimulus probabilities, a step increase in p(R) will result in an immediate movement of the starting point to a position close to the boundary for right responses, corresponding to a shift from SP_1 to SP_3 in figure 6. The starting point will then remain in this position until there is another change in p(R). This process gives rise to qualitative predictions for accuracy and response time which resemble those for Signal Detection Theory, while confidence predictions are the opposite of those for Signal Detection Theory. For example, since the starting point will lie close to the boundary for right responses after a step increase in p(R), many right responses will be made and, as a consequence, accuracy will be high.

Similarly, response time and confidence will vary with the position of the starting point. Since response time and confidence both vary as a direct function of the distance between the starting point and the boundary, times for correct right responses will be short and confidence will be low when the starting point lies close to the boundary for right responses following a step increase in p(R).

On the other hand, subjective probability may be determined on the basis of cumulative stimulus probabilities. In this case, as in Signal Detection Theory, the expected changes will be in the same direction as those described above, but the initial movement of the starting point will begin soon after the step change in p(R) and it will continue throughout the block. For example, a step increase in p(R) will give rise to a gradual movement of the starting point towards the boundary for right responses, corresponding to a gradual shift from SP_1 towards SP_3 in figure 6. As a consequence, there will be a ramp increase in the accuracy of right responses, while both times and confidence for right responses will show a ramp decrease.

(c) The accumulator model

In the accumulator model, changes in subjective probability will produce changes in the relative values of the criteria for the two responses. If subjective probability is determined on the basis of recent stimulus probabilities, a step increase in p(R) will produce an immediate decrease in the relative value of the criterion for right responses. Alternatively, if subjective probability is determined on the basis of cumulative stimulus probabilities, a step increase in p(R) will produce a gradual decrease in the relative value of the criterion for right responses. Therefore, the qualitative predictions of this process are identical to those described above for the random walk model.

2. The response stabilisation hypothesis

(a) Signal Detection Theory

When the response stabilisation hypothesis is applied to a Signal Detection Theory model, a mismatch in the local probability (L) and the cumulative probability (C) of a response will produce a movement in the cutoff. For example, when there is a step increase in p(R) from 0.2 to 0.8, the local probability of right responses, calculated over the last x trials, will increase to 0.8 within x trials. At the same time, if the practice trials described above were performed prior to the step increase in p(R), the cumulative probability of right responses will increase only slowly. Therefore, L will be greater than C, and the criterion will be adjusted upwards. In the experimental situation described above, the value of (L-C) for right responses will vary as shown in figure 44.

A computer simulation of the response stabilisation hypothesis applied to Signal Detection Theory was performed in the above experimental situation. Local probability was calculated when $x = 10$ trials. It was assumed that the initial value of the cutoff was 1.0 and the coefficient of adjustment of the criterion was set at 0.1. The simulated experiment was run 20 times and averaged. Figure 45(a) shows the expected values of β which correspond

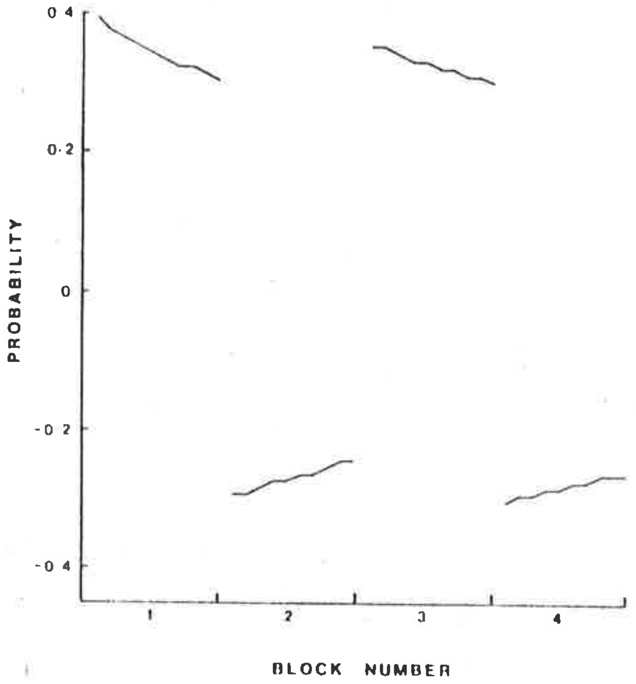


FIGURE 44. The expected changes in the difference between the local probability, L, and the cumulative probability, C, of a right response when the probability of a right stimulus, p(R), takes the values 0.8, 0.2, 0.8 and 0.2 in blocks of 100 trials which are numbered 1, 2, 3 and 4, respectively. It is assumed that 300 practice trials in which p(R)=0.5 and 100 practice trials in which p(R)=0.2 preceded the trials which are plotted.

to the values of (L-C) as shown in figure 44. As would be expected from the variation in (L-C), when there is a step increase in p(R), the expected value of β increases within x trials and continues to increase throughout the block.

The ramp increase in β following a step increase in p(R) corresponds to a gradual movement of the cutoff from x_{c_2} towards x_{c_1} in figure 41. From figure 41 it is apparent that, as the cutoff moves, there will be a progressive decrease in the probability that observations will fall on the side of the cutoff favouring right responses. Therefore, as β increases the accuracy of right responses will decrease. Figure 45(b) shows the variation in the probability of correct responses to right stimuli which corresponds to the

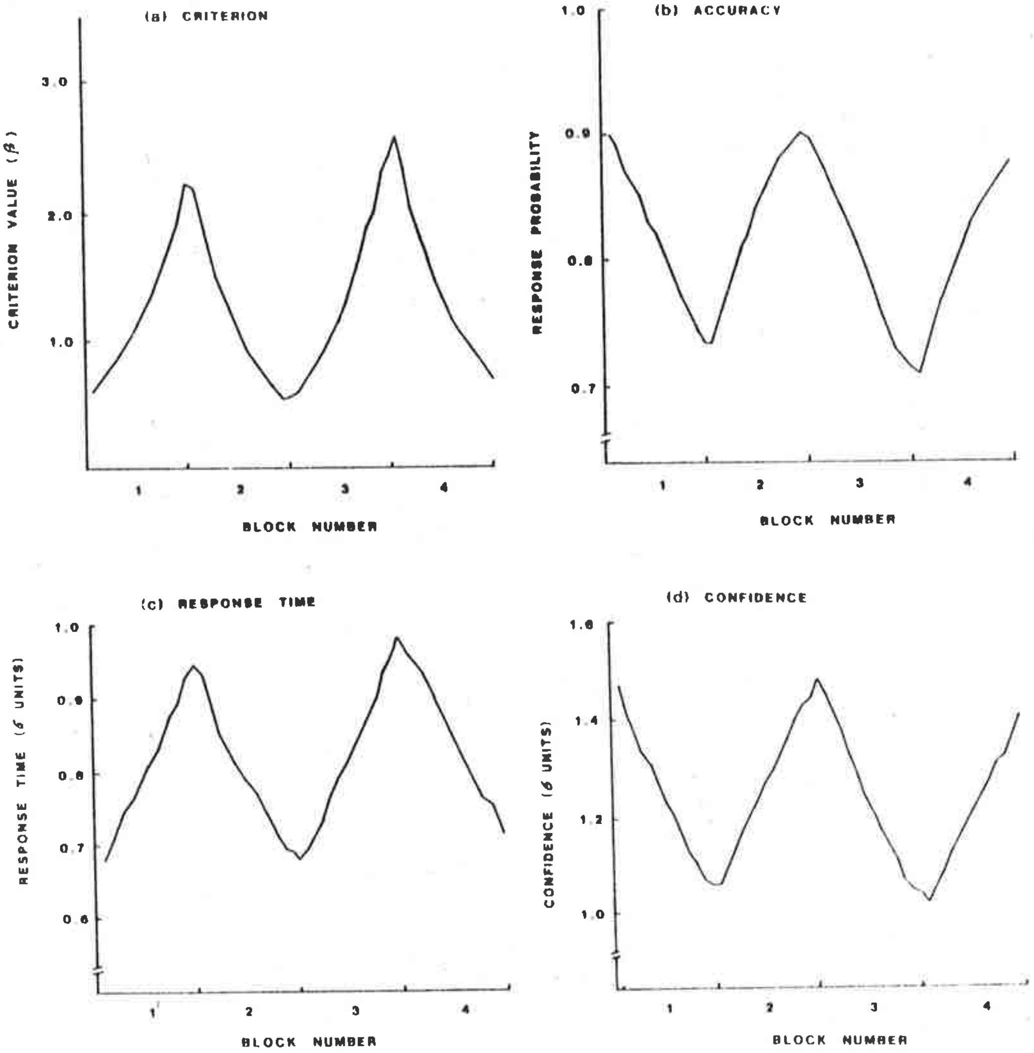


FIGURE 45. Pattern of change in (a) the criterion, β, (b) response probability, (c) response time and (d) confidence obtained in a simulation of the response stabilisation hypothesis applied to Signal Detection Theory when the probability of a right stimulus, p(R), takes the values 0.8, 0.2, 0.8 and 0.2 in blocks of 100 trials which are numbered 1, 2, 3 and 4, respectively. 300 practice trials in which p(R)=0.5 and 100 practice trials in which p(R)=0.2 preceded the trials which are plotted. Local probability was calculated over blocks of 10 trials. It was assumed that the initial value of β=1.0 and that the coefficient of adjustment of the criterion was 0.1. The distributions of sensory effect of left and right stimuli were assumed to be normal with means of 10 and 12, respectively, with σ=1.0 for both distributions and d'=2.0. Each curve is based on the average of 20 simulations of the experiment.

expected values of β shown in figure 45(a). Accuracy for right responses shows a ramp decrease in the block following a step increase in p(R).

Since response time varies as an inverse function of the distance between an observation and the cutoff, changes in the expected value of β will also give rise to changes in response time. The average distance between observations favouring right responses and the cutoff was estimated as described above. Response time will vary with the reciprocal of this distance as shown in figure 45(c). In other words, times for correct right responses show a ramp increase in the block following a step increase in p(R).

Confidence will also vary with β. Since confidence is a direct function of the distance between the observation and the cutoff, confidence for right responses will vary as shown in figure 45(d). A step increase in p(R) will produce a ramp decrease in confidence.

(b) The random walk model

In Chapter 1 it was noted that when the response stabilisation hypothesis is applied to the random walk model it determines the position of the starting point relative to the boundaries. For example, when L > C for right responses following a step increase in p(R) (see figure 44), the starting point will move away from the boundary for right responses. As the distance between the starting point and the boundary increases, fewer right responses will be made, so accuracy in right responses will decrease.

At the same time, response time and confidence will vary directly with the distance between the starting point and the boundary. Therefore, when L > C as a result of a step increase in p(R), times and confidence for right responses will increase as the starting point moves away from the boundary for right responses.

(c) The accumulator model

When the response stabilisation hypothesis is applied to the accumulator model, a mismatch in L and C for a given response will produce a change in the value of the criterion for that response. For example, when $L > C$ following a step increase in p(R), the criterion for right responses will increase. The resulting qualitative predictions for accuracy, response time and confidence are the same as those described above for the random walk model.

3. **The adaptation-level hypothesis**

(a) Signal Detection Theory

When the adaptation-level hypothesis is applied to Signal Detection Theory, the position of the cutoff is directly determined by the adaptation level. Since a change in p(R) produces a change in the value of the adaptation level, it will give rise to a movement of the cutoff. For example, figure 46(a) shows the actual changes in the adaptation level, calculated when $x = 10$ trials, in the experimental situation described above. The adaptation level shows a ramp increase following a step increase in p(R). In addition, it can be seen that the magnitude of the variation in the adaptation level decreases with each successive step change in p(R). Since accuracy, response time and confidence vary as a function of the adaptation level, there will be a corresponding decrease in the amount of variation in these measures with successive changes in p(R).

A ramp increase in the adaptation level following a step increase in p(R) corresponds to the gradual movement of the cutoff from x_{c_2} towards x_{c_1} in figure 41. As the cutoff moves, the probability that observations will favour right responses decreases, so accuracy will decrease. Figure 46(b) shows the variation in the probability of correct right responses which

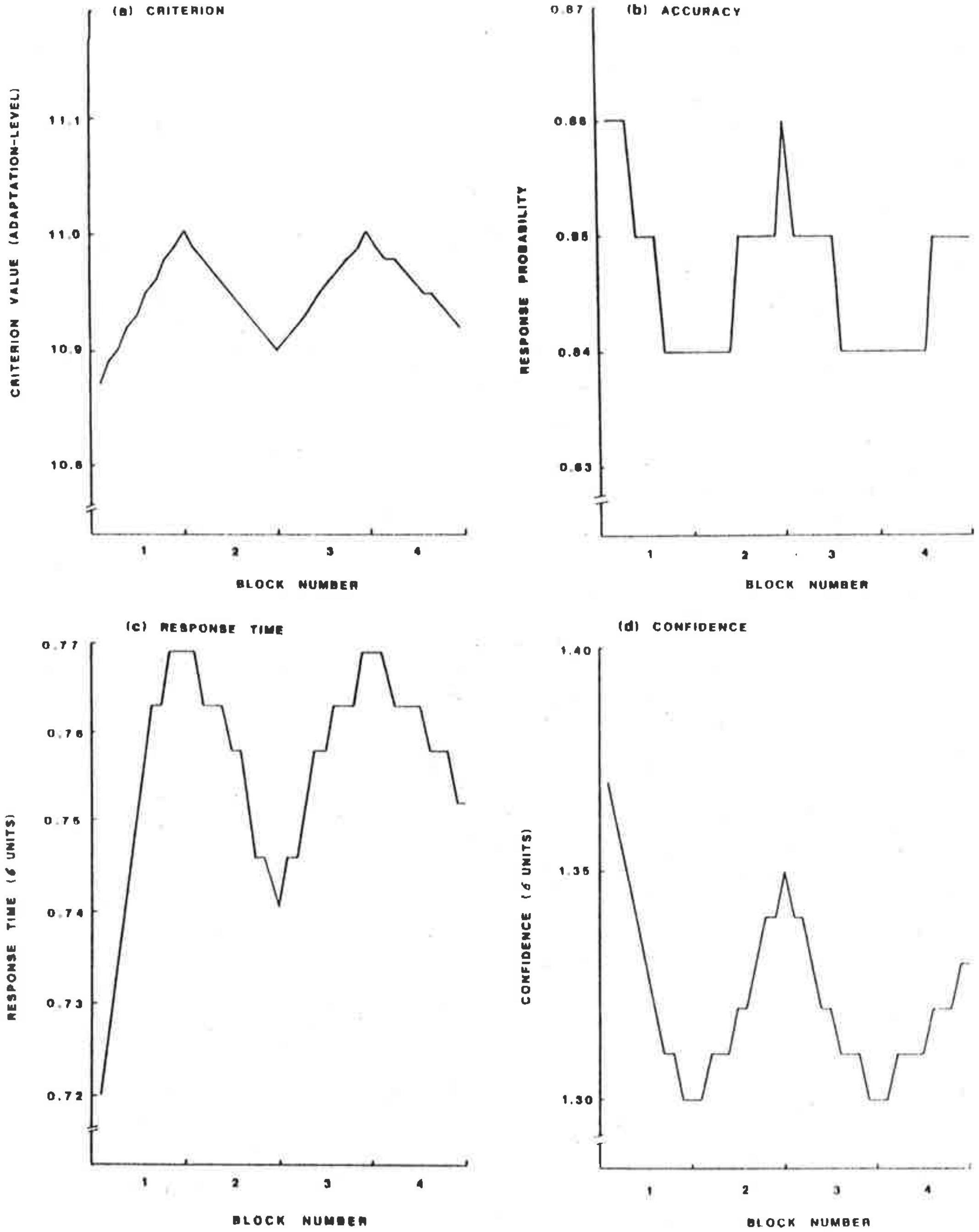


FIGURE 46. Predictions of the adaptation-level hypothesis applied to Signal Detection Theory when the probability of a right stimulus, p(R), takes the values 0.8, 0.2, 0.8 and 0.2 in blocks of 100 trials which are numbered 1, 2, 3 and 4, respectively. It is assumed that 300 practice trials in which p(R)=0.5 and 100 practice trials in which p(R)=0.2 preceded the trials which are plotted. In addition, it is assumed that the distributions of the sensory effect of left and right stimuli are normal with means of 10 and 12, respectively, with σ=1.0 for both distributions and d'=2.0. The figures show the expected variation in (a) the criterion, β, (b) response probability, (c) response time and (d) confidence.

corresponds to the expected values of the cutoff in figure 46(a). After a step increase in p(R) there is a ramp decrease in accuracy for right responses.

Since response time varies as an inverse function of the distance between an observation and the cutoff, the changes in the position of the cutoff as determined by the adaptation-level hypothesis will produce changes in response time. For each value of the cutoff in figure 46(a), the average distance between an observation favouring right responses and the cutoff was estimated as described above. Times for correct right responses will vary with the reciprocal of this distance, as shown in figure 46(c). There is a ramp increase in times for right responses following a step increase in p(R).

Since confidence is a direct function of the distance between an observation and the cutoff, it will vary as shown in figure 46(d). There is a ramp decrease in confidence in right responses following a step increase in p(R).

(b) The random walk model

As described in Chapter 2, in the random walk model it may be assumed that the indifference point, which separates those observations which favour one response from those which favour the alternative, directly reflects the adaptation level. In this case, when there is a step increase in p(R), the indifference point will increase to take on a value closer to the intensity of right stimuli. As a consequence, the effective discriminability of right stimuli will be reduced, so fewer right responses will be made and their accuracy will decrease.

Since many observations will be needed before a right response is made, times for right responses will increase following a step increase in p(R). At the same time, since the effective discriminability of right stimuli is reduced, following Ascher (1974) (see equation (7)) confidence in right responses will decrease.

(c) The accumulator model

The adaptation-level hypothesis can also be assumed to determine the value of the indifference point in the accumulator model. For example, a step increase in p(R) will give rise to an increase in the value of the indifference point with the result that right stimuli will be less discriminable. Therefore, the qualitative predictions of this process are identical to those described above for the random walk model.

4. The target confidence mechanism

In Chapter 1 it was noted that it is not possible to combine the mechanism based on target confidence with Signal Detection Theory, so discussion is restricted to the variable sample models.

(a) The accumulator model

In the accumulator model, the target confidence mechanism controls the values of the criteria in the primary decision process. For example, if it is assumed that the target confidence values for both responses are the same, a step increase in p(R) will give rise to an increase in the proportion of trials on which right responses are made correctly and with high confidence. As a result, on the majority of trials, actual confidence in right responses will tend to exceed target confidence, so overconfidence will be stored in the appropriate control accumulator. Therefore, the critical amount of overconfidence is likely to be reached before the criterion for underconfidence is satisfied, and the criterion for right responses in the primary decision process will be lowered.

Changes in the relative values of the criteria in the primary decision process will give rise to variations in the response measures. For example, the decrease in the relative value of the criterion for right responses following a step increase in p(R) will result in an increase in the number of right responses being made, with a consequent increase in accuracy.

Since the criterion for right responses is low after a step increase in p(R), few observations will be needed to satisfy it, so times for right responses will be short. In addition, when the criterion for right responses is reduced following a step increase in p(R), the maximum possible confidence in right responses will be equal to the lower criterion value. Therefore, confidence in right responses will tend to be low.

These predictions can be confirmed by comparison with a computer simulation of the accumulator model performing a closely similar signal detection task which was carried out by Vickers (1979). The major difference between the signal detection process and the discrimination situation is that, in the signal detection process the criteria for signal and nonsignal responses are not symmetrical, while in the discrimination process the criteria for the two responses may be assumed to be identical. In particular, in this simulation the specification of signal and nonsignal criteria incorporates a bias towards signal responses. However, this is the only difference between the two forms of the model which is important in this situation. In the simulation it was assumed that the target levels of confidence for signal responses and nonsignal responses were equal at the start of the trials with a value of 6.0, and that the criteria for both responses in the primary decision process were equal to 5.0. The difference between the means of the distributions of sensory effect for signal and noise was set at 1.80 and the coefficient of adjustment of the criteria was 1.50. Twenty blocks each of 50 trials were run. In the first six blocks the *a priori* probability of a signal was 0.5. Following this, the probability of a signal was 0.2 for two blocks and then 0.8 for two blocks and it continued to alternate in this way for the remaining blocks of trials. Therefore, the manipulation of the signal probability was equivalent to the manipulation of p(R) in the experimental situation described above, and the results for signal responses may be used to indicate the expected variation in right responses. The simulated experiment was run 10 times and the data were averaged.

Figure 47(a) shows the variation in the probability of correct signal responses (hits) which results from step changes in signal probability. Following a step increase in the probability of a signal, the accuracy of signal responses is high. Figure 47(a) also shows the variation in the probability of correct nonsignal responses (correct rejections). It is apparent that the probability of correct rejections varies in the complementary way to the probability of hits. However, the probability of hits is always higher than the probability of correct rejections. This is the result of the bias towards signal responses which is incorporated in the specification of the criteria, and therefore it is not a necessary prediction of the model. In addition, from the figure it can be seen that there is no consistent pattern in the variation in accuracy within probability conditions, even though, in the majority of cases there appears to be a decrease in accuracy from the first to the second half of blocks. These trends are an indication of the state of the adaptive process at the time when each step change in probability takes place, and result from the imposition of a cyclic step change in probability on a periodic adaptive process which is changing with a different periodicity. Since previous simulations of the accumulator model have shown that the initial response to a step change in probability occurs within 10 trials (e.g. Vickers, 1979, fig. 143), the unsystematic variation of accuracy within blocks is not critical to the predictions of the model.

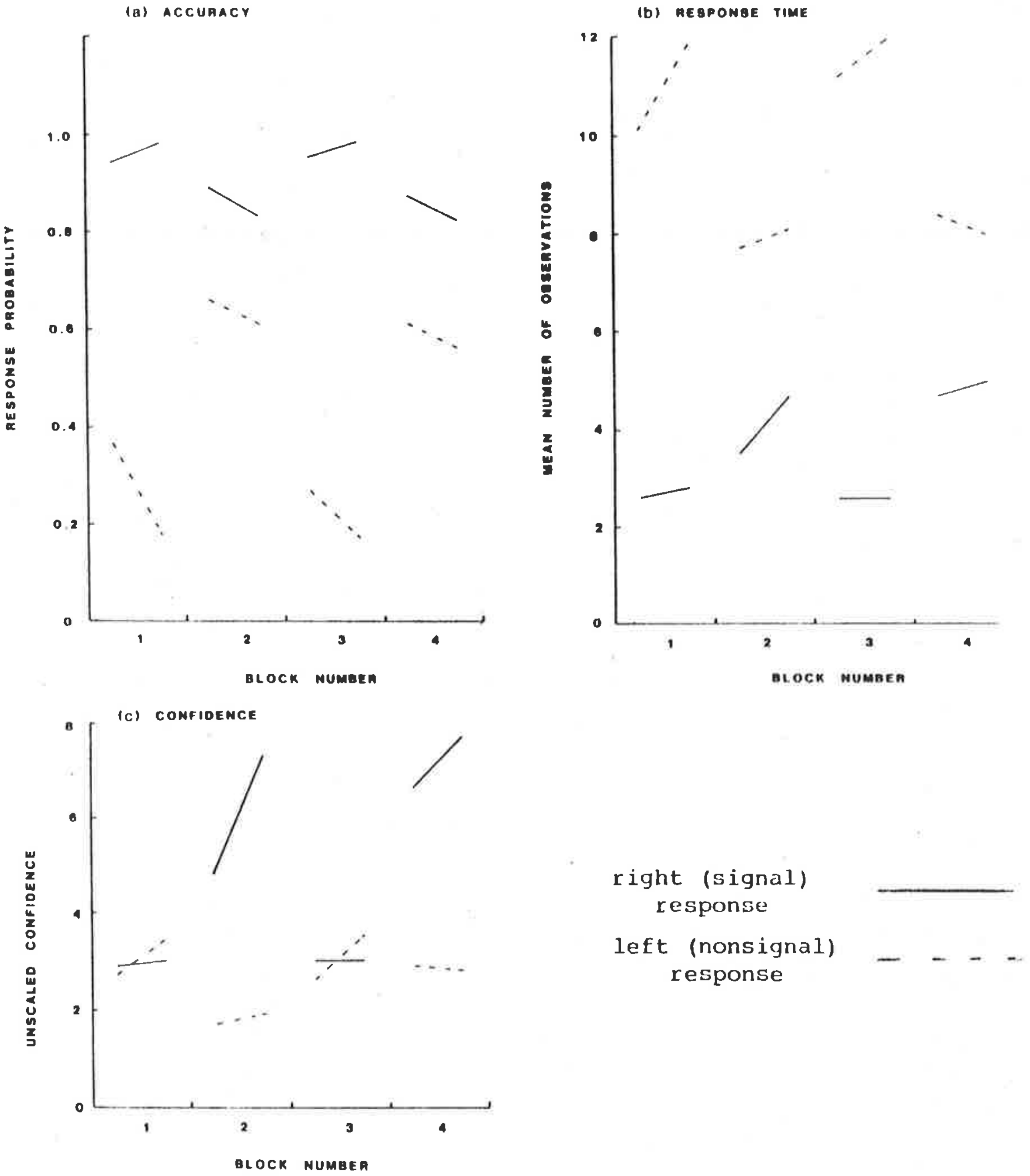


FIGURE 47. Patterns of change in (a) response probability, (b) response time and (c) confidence, obtained in a simulation of the target confidence mechanism applied to the accumulator model performing a signal detection task which was carried out by Vickers (1979). The specification of signal and nonsignal criteria incorporated a bias towards signal responses. It was assumed that target confidence for signal (right) responses and nonsignal responses was equal at the start of the trials, with a value of 6.0 and that the criteria for both responses in the primary decision process were equal, with a value of 5.0. The difference between the means of the distributions of sensory intensities of signal and noise was set at 1.80 and the coefficient of adjustment of the criteria was 1.50. 300 practice trials in which p(S)=0.5 and 100 practice trials in which p(S)=0.2 preceded the trials which are plotted. The probability of a signal was 0.8, 0.2, 0.8 and 0.2 in blocks of 100 trials which are numbered 1, 2, 3 and 4, respectively. Mean values are shown for the first and second half of each block.

Figure 47(b) shows the pattern of response times for hits and correct rejections from the simulation. As with accuracy measures, there is clear evidence of a bias towards signal responses which are always made more quickly. However, times for hits and correct rejections show the same pattern of variation. In particular, times for hits are short following a step increase in probability. Although the variations in response time within blocks of trials appear to be systematic, the pattern of change is not a specific prediction of the model. Instead, as for accuracy, the changes in response time within blocks indicate the state of the adaptive decision process when each step change in probability took place.

The variation in confidence in hits and correct rejections obtained in the simulation is shown in figure 47(c). From the figure it can be seen that confidence in hits is low after a step increase in probability. Confidence in correct rejections varies in the complementary way. However, the bias towards signal responses in the specification of the criteria is evident in blocks 2 and 4 where confidence in hits is much higher than confidence in correct rejections. In addition, as for accuracy and response time, the variations in confidence within blocks indicate the state of the adaptive decision process at the time of each step change in probability and therefore, these changes do not correspond to a specific prediction of the model.

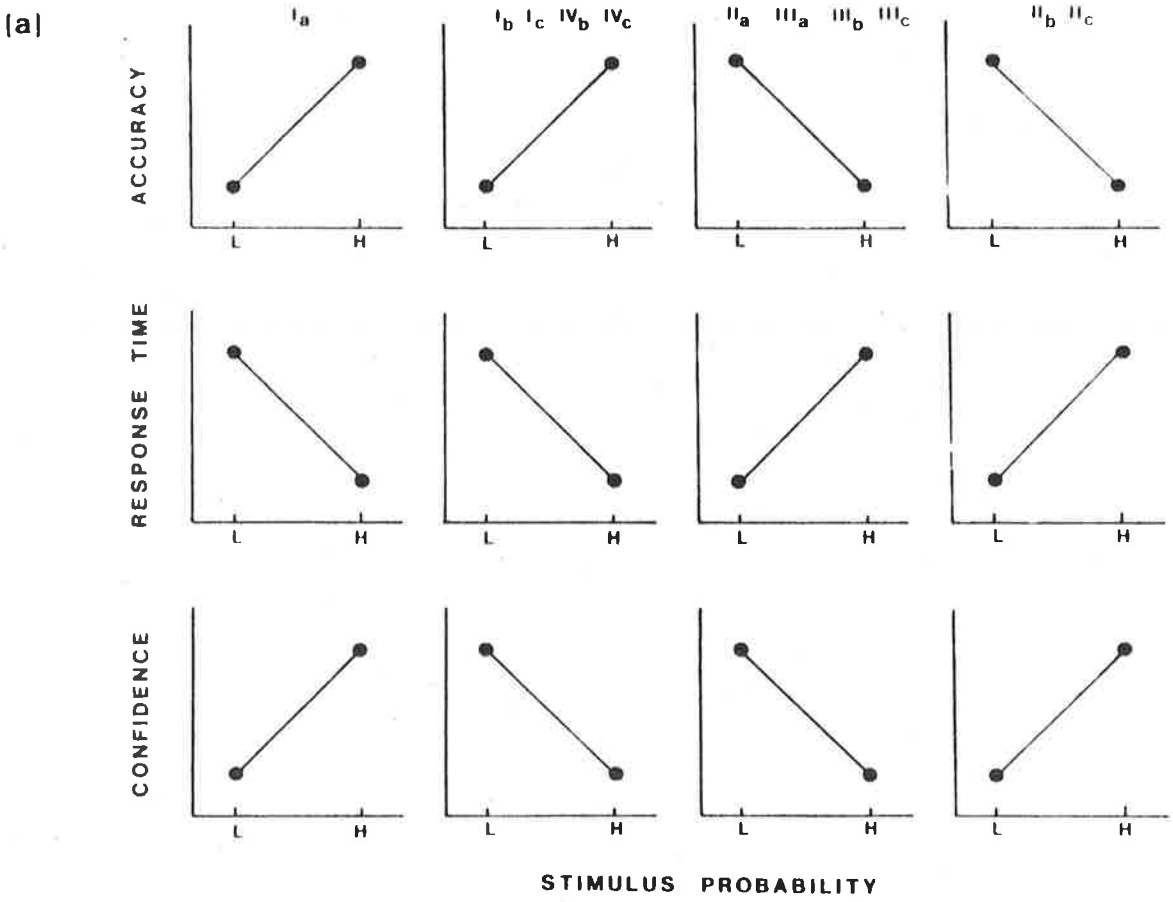
(b) The random walk model

As discussed in Chapter 2, in order to be able to apply the target confidence mechanism to the random walk model it is necessary to assume that the value of the desired likelihood ratio diminishes over time. In other words, the distance between both of the boundaries and the starting point decreases with each successive observation taken before a decision is made (see figure 8). Within this model, the target confidence mechanism will determine the position of the starting point relative to the diminishing boundaries. For example, if it is assumed that the target confidence values for both responses are the same, a step increase in p(R) will result in a decrease in the number of right responses made incorrectly and with low confidence. As a result, the actual confidence in right responses will tend to exceed target confidence. Therefore, the starting point will be likely to move towards the boundary for right responses. This process is analogous to the target confidence mechanism applied to the accumulator model and its predictions are qualitatively identical to those described above for the accumulator model.

5. Summary

The qualitative predictions of the mechanisms for criterion regulation applied to each decision model are summarised in figure 48. All of the predictions are given in terms of right responses, so a high value of stimulus probability, H, indicates that there has been a step increase in p(R), and a low value of stimulus probability, L, indicates that there has been a step decrease in p(R). From the figure the processes appear to be grouped on the basis of the mechanisms for criterion control and their application to specific decision models.

In particular, the predictions of the ideal observer hypothesis and the target confidence mechanism appear to contrast with the predictions of the response stabilisation and adaptation-level hypotheses. When either the ideal observer hypothesis or the target confidence mechanism is applied to any of the decision models, the accuracy and response time predictions of the resulting processes are the same. Similarly, when either the response stabilisation hypothesis or the adaptation-level hypothesis is applied to any of the decision models, the accuracy and response time predictions of the



(b)

CRITERION REGULATION MECHANISM	DECISION MODEL		
	(a) Signal Detection Theory	(b) random walk model	(c) accumulator model
I: ideal observer hypothesis	Ia	Ib	Ic
II: response stabilisation hypothesis	IIa	IIb	IIc
III: adaptation-level hypothesis	IIIa	IIIb	IIIc
IV: target confidence mechanism		IVb	IVc

FIGURE 48. A summary of the qualitative predictions of each of the mechanisms for criterion regulation applied to each of the decision models when p(R) varies between 0.2 and 0.8 in step changes between successive blocks. All of the predictions are given in terms of responses to right stimuli. Stimulus probability L indicates a block in which p(R)=0.2 while stimulus probability H indicates a block in which p(R)=0.8. In figure 48(a) the decision processes are labelled as defined in figure 48(b).

resulting processes are the same. However, in this case, the predicted changes in accuracy and response time are the opposite of those predicted by the ideal observer hypothesis and the target confidence mechanism. Therefore, it appears that accuracy and response time predictions depend on the form of the mechanism for criterion regulation.

In contrast, the confidence predictions of the processes appear to depend on the nature of the decision model. For the ideal observer hypothesis and the target confidence mechanism, the critical factor appears to be whether a decision is based on a single observation, as in Signal Detection Theory, or on multiple observations, as in the random walk and accumulator models. From figure 48 it is apparent that, for this pair of criterion regulation mechanisms, the confidence predictions for Signal Detection Theory are the opposite of the confidence predictions for either the random walk model or the accumulator model. On the other hand, this distinction does not hold for the response stabilisation and adaptation-level hypotheses. In this case, when either of these mechanisms is applied to Signal Detection Theory or when the adaptation-level hypothesis is applied to the random walk or accumulator models, confidence predictions are the opposite of those obtained when the response stabilisation hypothesis is applied to the random walk or accumulator models.

This categorisation highlights the interaction between the nature of the mechanism for criterion regulation and the type of decision model, when considering predictions for accuracy, response time and confidence. Consideration of the predictions for accuracy and response time results in a dichotomous classification of the processes in terms of the mechanisms for criterion regulation. An examination of confidence predictions then distinguishes between the processes within each pairing on the basis of the combination of criterion regulation mechanisms with specific decision models. Therefore, it appears that the measurement of confidence is critical to any comparison of the processes.

B. EXPERIMENT 5

1. Introduction

In order to evaluate the predictions of the different mechanisms for criterion regulation applied to any of the three decision models, measures of accuracy, response time and confidence were taken in a situation which allowed for a more detailed examination of the data than was possible in Experiments 3 and 4. Of particular interest was the nature of the variation in accuracy and in response time and the relationship between these changes and the pattern of confidence ratings. The experimental design was the same as in Experiment 4 except that discriminability was held constant in order to reduce the variability within bias conditions. Since the order in which observers performed the two bias conditions did not account for a significant proportion of the variance in Experiment 4, in this experiment all observers performed under the same order condition. As a consequence, the effective sample size was increased, producing an increase in the reliability of response measures.

2. Method

(a) Stimuli

The stimuli were white pi-shaped figures identical to those used in Experiment 3 and described in Chapter 4, section A.2(a) except that, in this case, discriminability did not vary. The longer line was always 187 mm in length while the shorter line was always 186 mm in length.

(b) Apparatus

The apparatus was identical to that used in Experiment 3 and described in Chapter 4, section A.2(b).

(c) Observers

The 30 observers, 15 men and 15 women, were students enrolled in the first year psychology course at the University of Adelaide, whose participation was credited towards a course requirement. They ranged in age from 17 to 24 years and all declared themselves to be right-handed. All observers were naive with respect to the aims of the experiment.

(d) Design

All observers performed under the same conditions with each observer being tested individually for approximately one hour.

Stimuli were presented in 10 blocks: 4 practice blocks each of 50 trials preceded 6 experimental blocks each of 100 trials. In each practice block and the first experimental block, the occurrence of the longer line on the right or the left was equiprobable ($p(R) = 0.5$). The remaining five blocks of experimental trials took one of two forms:

A. The probability of the longer line occurring on the right was 0.8 and therefore, the probability of the longer line occurring on the left was 0.2.

B. The probability of the longer line occurring on the right was 0.2 and therefore, the probability of the longer line occurring on the left was 0.8.

All observers performed the last five experimental blocks in the order BABAB. The first two experimental blocks were treated as practice, so only the last four experimental blocks were considered in the analyses.

Within each block, both practice and experimental, the occurrence of the longer line on the right or the left was random, given the probability restriction.

Consecutive practice blocks were separated by short rest periods in which observers were given the opportunity to clarify instructions. The six experimental blocks were performed in an uninterrupted sequence, and observers were not given any indication of the manipulation of a priori probability.

(e) Procedure

The procedure was identical to that of Experiment 3 and described in Chapter 4, section A.2(e).

3. Results

Analyses of variance with Period (first two experimental blocks or second two experimental blocks), Probability (block type A or type B as defined in section B.2(d)), Half (first or second half of an experimental block) and Response (right or left) as factors were used to analyse data from the last four blocks of experimental trials. Analyses of variance were carried out according to a crossed factorial design with repeated measures on all factors. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses.

(a) Accuracy

The analysis of accuracy data showed a significant interaction between Probability, Half and Response ($F(1,29) = 41.18$, $MSerror = 0.13$, $p < .001$)

The percentages of right and left stimuli for which responses were correct in each half of each type of block are shown in table 16. Halves 1 and 2 of Probability A in which $p(R) = 0.8$ are labelled H1 and H2, respectively, and Halves 1 and 2 of Probability B in which $p(R) = 0.2$ are labelled L1 and L2, respectively. When there was a step increase in $p(R)$ from 0.2 to 0.8 between L2 and H1, there was a significant increase in the accuracy of responses to right stimuli and a significant decrease in the accuracy of

RESPONSE MEASURE		PROBABILITY			
		L2	H1	H2	L1
ACCURACY	Right	88.51	92.03	85.30	82.51
	Left	85.81	83.04	91.00	90.04
RESPONSE TIME	Right	1016.99	977.28	1032.68	1051.17
	Left	1064.39	1111.61	1037.10	1033.04
CONFIDENCE	Right	3.04	3.17	2.91	2.87
	Left	2.72	2.70	2.81	2.93

TABLE 16. Percentage of correct responses, mean response times and mean confidence for correct right and left responses in Half 1 (H1) and Half 2 (H2) of blocks of Probability A and in Half 1 (L1) and Half 2 (L2) of blocks of Probability B. The measures were taken over 30 observers.

responses to left stimuli. On the other hand, when there was a step decrease in p(R) from 0.8 to 0.2 between H2 and L1, there was a significant decrease in the accuracy of responses to right stimuli and a nonsignificant decrease in the accuracy of responses to left stimuli. At the same time, accuracy for right responses decreased significantly and accuracy for left responses increased significantly from the first to the second half of blocks of type A (i.e. from H1 to H2). In contrast, accuracy for right responses increased significantly and accuracy for left responses decreased significantly from the first to the second half of blocks of type B (i.e. from L1 to L2).

(b) Response time

The analysis of response time data showed a significant main effect due to Response (F(1,29) = 9.58, MSerror = 22099, $p < .01$) and significant interactions of Period x Probability (F(1,29) = 8.41, MSerror = 6247, $p < .01$), Probability x Response ($F(1,29) = 6.19$, MSerror = 14513, $p < .05$) and Probability x Half x Response ($F(1,29) = 19.14$, MSerror = 14966, $p < .001$).

The Response main effect arose because right responses were faster than left responses. The mean times for right and left responses were 1019.53 msec and 1061.54 msec, respectively.

The Period x Probability interaction reached significance because, in Probability A, the mean times for responses in Period 1 were significantly longer than the mean times for responses in Period 2. The respective mean response times were 1060.39 msec and 1018.94 msec. In contrast, there was no difference between the times for responses in Periods 1 and 2 in Probability B. The mean times for responses in blocks of type B in Periods 1 and 2 were 1041.19 msec and 1041.60 msec, respectively.

The Probability x Response interaction is shown in table 17 where it can be seen that right responses were faster in Probability A than in Probability B. Conversely, left responses were faster in Probability B than in Probability A.

	PROBABILITY	
RESPONSE	A	B
Right	1004.98	1034.08
Left	1074.34	1045.71

TABLE 17. Mean times for correct right responses and correct left responses in blocks of Probability A and of Probability B. The means were taken over 30 observers.

The Probability x Half x Response interaction is summarised in table 16 where it can be seen that, when there was a step increase in p(R) between L2 and H1, there was a significant decrease in the times for right responses and a significant increase in the times for left responses. Conversely, when there was a step decrease in p(R) between H2 and L1, there was a nonsignificant increase in the times for right responses and a nonsignificant decrease in the times for left responses. In addition, times for right responses increased significantly and times for left responses decreased significantly from Half 1 to Half 2 of blocks of type A (i.e. from H1 to H2). In contrast, times for right responses showed a nonsignificant decrease and times for left responses showed a nonsignificant increase from Half 1 to Half 2 of blocks of type B (i.e. from L1 to L2).

(c) Confidence

The analysis of confidence data showed a significant main effect due to Response (F(1,29) = 8.47, MSerror = 0.62, $p < .01$) and significant interactions of Probability x Response (F(1,29) = 4.60, MSerror = 0.17, $p < .05$) and Probability x Half x Response (F(1,29) = 31.62, MSerror = 0.13, $p < .001$).

The Response main effect arose because confidence was higher in right responses than in left responses. The mean confidence for right and left responses was 3.00 and 2.79, respectively.

The Probability x Response interaction is shown in table 18 where it can be seen that confidence in right responses was higher in Probability A than in Probability B. Conversely, confidence in left responses was higher in Probability B than in Probability A.

The Probability x Half x Response interaction is summarised in table 16. When there was a step increase in p(R) between L2 and H1 there

RESPONSE	PROBABILITY A	PROBABILITY B
Right	3.04	2.95
Left	2.75	2.82

TABLE 18. Mean confidence in correct right responses and correct left responses in blocks of Probability A and of Probability B. The means were taken over 30 observers.

was a significant increase in confidence in right responses and a nonsignificant decrease in confidence in left responses. When there was a step decrease in p(R) between H2 and L1 there was a nonsignificant decrease in confidence in right responses and a significant increase in confidence in left responses. In addition, confidence in right responses decreased significantly and confidence in left responses increased significantly from Half 1 to Half 2 of blocks of type A (i.e. from H1 to H2). Conversely, there was a significant increase in confidence in right responses and a significant decrease in confidence in left responses from Half 1 to Half 2 of blocks of type B (i.e. from L1 to L2).

(d) Confidence and response time

For each observer, confidence in all responses (correct and incorrect) was plotted against the corresponding response time values and straight lines were fitted. Separate plots were made for right and left responses in each block. It was found that 146 of the total of 240 plots had a negative slope, showing that overall, there was an inverse relationship between confidence and response time within blocks ($p < .001$, binomial test). The predominantly inverse relationship was still evident when the 60 plots for each of right and left responses were considered separately in each type of block. In blocks of type A, the numbers of negative slopes for right and left responses were 43 and 39, respectively ($p < .001$, $p < .05$, respectively, binomial test). In blocks of type B, the numbers of negative slopes for right and left responses were 37 and 44, respectively ($p < .05$, $p < .001$, respectively, binomial test). In all cases, the majority of the best fitting straight lines had a negative slope indicating an inverse relationship between confidence and response time for responses to both the less and the more probable stimulus in each probability condition.

4. Discussion

Several of the effects which reached significance do not appear to have arisen from the manipulation of bias by varying a priori probability. For example, both response time and confidence measures showed a significant main effect of Response. Right responses were faster and more confident than left responses. This appears to result from a general, pre-existing bias towards right responses shown by right-handed observers. Although this finding does not correspond to a specific prediction of any of the processes, it is not inconsistent with any of them.

Similarly, the Period x Probability interaction in response time data does not appear to reflect the manipulation of response bias. Instead, this interaction may be interpreted in terms of a mismatch between the observer's expectations or probability assessment and his responding, combined with the influence of a general bias towards right responses. The mean response time in blocks of type B, in which most responses were left responses, did not change from Period 1 to Period 2. In contrast, in the block of type A in Period 1, which was the first experimental block, the mean response time was longer than in any other block. It seems likely that, on the basis of the initial practice trials, observers would expect that the stimuli would be equiprobable. The discrepancy between this assessment and the evidence from their responses in the last practice block and the first experimental block (the block of type A in Period 1) may have made observers more cautious in these blocks and therefore, response times would be longer. It appears that the manipulation of a priori probability may have become apparent by the block of type B in Period 1, producing a change in expectations to match responding, with a consequent decrease in response times. The faster times in the block of type A in Period 2 appear to reflect the bias towards right responses of right-handed observers since, in this block, $p(R) = 0.8$.

The effects of the step changes in a priori probability were evident in the Probability x Half x Response interaction which accounted for a significant proportion of the variance in all three response measures. In addition, the Probability x Response interaction was significant for response time and confidence. These effects will be discussed in terms of the higher order interaction. The Probability x Half x Response interaction comprises two major components, with each response measure showing variations both between probability conditions and within probability conditions. These will be discussed in turn.

The changes in accuracy and response time which occurred as the result of step changes in probability are consistent with Laming's (1968, 1969) findings and conform to the predictions of the ideal observer hypothesis or the target confidence mechanism applied to any of the decision models. However, these results run contrary to the predictions of the response stabilisation or adaptation-level hypotheses applied to any of the models.

When confidence ratings are considered, it is possible to make a further distinction between the ideal observer hypothesis and the target confidence mechanism and their application to the three decision models. In particular, the pattern of change observed in confidence is not consistent with the target confidence mechanism as applied to any of the models. Confidence findings also contrast with the adjustment of criteria to conform to changes in subjective probability in either the random walk or accumulator models. However, when the ideal observer hypothesis is applied to Signal Detection Theory, the process predicts the observed variation in confidence as well as the changes in accuracy and response time. In addition, when confidence data are considered alone, they are consistent with the response stabilisation hypothesis applied to the random walk model or the accumulator model. However, as seen above, these processes cannot account for the accuracy and response time data.

In the introduction to this chapter it was pointed out that an important feature which distinguishes between the different decision processes is the nature of the predictions for confidence. While the changes in accuracy and response time between probability conditions in the present experiment are consistent with a range of processes, on the other hand, consideration of the confidence data makes it possible to distinguish between them. Since only Signal Detection Theory predicts the findings for all three response measures, it appears that there may be a need for some modification of the formulations for confidence in the random walk and accumulator models. However, since these models can account for confidence data when criteria are regulated by the response stabilisation hypothesis, it may be that it is the nature of the criterion regulation mechanism, which indirectly determines confidence, which is inappropriate rather than the formulation for confidence itself.

Empirical support for the ideal observer hypothesis applied to Signal Detection Theory has also been obtained in the closely related field of vigilance performance. In particular, in studies in which signal probability underwent a step change between the practice and experimental trials but remained constant thereafter (e.g. Colquhoun, 1961; Colquhoun & Baddeley, 1964, 1967), or when instructions indicated the actual signal probability (e.g. Williges, 1969, 1971, 1973), observers adopted optimal values of β as determined by the ideal observer hypothesis.

However, a contradiction appears to arise from several vigilance studies in which there were apparent or actual changes in a priori probability during the experimental session. For example, by giving observers the misleading expectation that signal probability would be 0.5, Williges (1969) induced an apparent gradual change in signal probability during the experimental session of a detection experiment. In this situation, Williges (1969) found that observers did not adopt optimal values for β as determined by the ideal observer hypothesis on the basis of either the expected signal probability or the actual signal probability. Instead, in contrast to the ideal observer hypothesis, observers adopted a high value for β when the actual signal probability was high and a low value for β when the actual signal probability was low. Similar results were obtained in a detection experiment by Vickers and Leary (1983) in which there was a ramp decrease in signal probability from 0.5 to 0.06 during the experimental session. In this case, observers adopted β values which decreased with signal probability. In a similar discrimination experiment, Vickers et al. (1977) found that when the probability of one stimulus underwent a ramp decrease from 0.5 to 0.05 during the session, observers also decreased their criterion values.

One possible reason for the mismatch between these sets of findings appears to be the different ways in which a priori probability was varied in these studies. In both the present experiment and the previous detection studies which found support for the ideal observer hypothesis applied to Signal Detection Theory, the changes in a priori probability were obvious to observers. Either a priori probability was known exactly or it underwent large, abrupt changes which would have been apparent even to observers who were not told to expect these changes. In contrast, in Williges' (1969) study the difference between the expected signal probability and the actual signal probability would have become apparent only gradually, if at all, as the trials progressed. Similarly, in the studies of Vickers and Leary (1983) and Vickers et al. (1977) a priori probability changed very slowly. Therefore, it appears that the knowledge or expectation that a priori probability will vary, results in observers selecting criteria as determined by the ideal observer hypothesis. This is consistent with Vickers' (1979) suggestion that there are two indicators of the stimulus situation operating on the decision mechanism, one of which represents the actual stimulus situation and the other of which incorporates expectations about it. Further evidence for the existence of two such indicators is found when the changes within and between probability conditions in the current data are considered together, so this notion will be discussed below with reference to these findings.

Although the changes in the response measures which occur between probability conditions are well described by the ideal observer hypothesis applied to Signal Detection Theory, the changes within probability conditions are less easily explained. Variations in the response measures within blocks of trials are not always a specific prediction of the decision processes. However, evidence of any change within blocks has some implications for the form of the process which is operating. In particular, the finding that response measures varied within blocks is inconsistent with the ideal observer hypothesis based on recent stimulus probability. In addition, from table 16 it is apparent that the changes within blocks did not correspond to a continuation of the changes between blocks since they were in the opposite direction. Therefore, these data also conflict with the ideal observer hypothesis based on cumulative stimulus probability according to which the initial change due to a step change in probability will continue smoothly throughout the subsequent block.

When the variations in the three response measures which occurred within and between blocks are considered together it is apparent that, in each case, the initial change in response to a step change in probability underwent a reversal during the subsequent block. One possible explanation for this is that there are two indicators of the stimulus situation which are varying with different periodicities. For example, Vickers and Leary (1983) have suggested that the conflict between the results of traditional vigilance studies and the studies of Vickers et al. (1977) and Vickers and Leary (1983) can be resolved if criteria are determined on the basis of either the response stabilisation hypothesis or the adaptation-level hypothesis, both of which incorporate two indicators of the stimulus events. In the response stabilisation hypothesis, criteria are determined on the basis of the values of L and C. Alternatively, according to the adaptation-level hypothesis observers adopt a criterion equal to their adaptation level. It is possible that, initially, the criterion might be based on a measure of recent probability or on expectations. The subsequent adaptation level could be viewed as the effect of cumulative changes on this initial level. However, as seen above, the decrease in accuracy and confidence and the increase in response time following a step increase in p(R) which are predicted by these two mechanisms, contrast with the present data.

In addition, an examination of the rate at which the two indicators of the stimulus events vary in each of these mechanisms shows that the predicted variations within blocks of trials will always be in the same direction as the initial change, so that the changes in response measures would not be expected to undergo a reversal during the block. For example, in the response stabilisation hypothesis both L and C are directly related to the stimulus probability, so both L and C will always vary in the same direction in this type of experimental situation. In particular, when there is a step increase in p(R), the local probability of right responses will increase immediately to take a value near p(R). At the same time, the cumulative probability of right responses will immediately start to increase, but at a slower rate. Therefore, both L and C will increase throughout a block and so, the initial change following a step increase in p(R) will continue throughout the block (see figure 44). The direction of change in both L and C will be reversed only when there is a subsequent step decrease in p(R). As a consequence, L and C have the same periodicity in this situation and do not give rise to reversals in performance. The indicators in the adaptation-level hypothesis will also vary in the same direction and with the same periodicity in this experimental situation. Indeed, from the predictions of the processes set out in the introduction it can be seen that the only process which shows any evidence of reversals in performance is the target confidence mechanism applied to the accumulator model. However, as described in the introduction, this is not a specific prediction of the process. Instead, it results from the state of the adaptive process at the time of each step change in probability. As a consequence, there may be no reason to expect that the changes within blocks will always be in the opposite direction to the changes between blocks.

It appears that none of the existing processes incorporates independent indicators of the stimulus situation which, in combination, could give rise to the observed variation within and between blocks. However, these data are not inconsistent with Vickers' (1979) notion that there is one indicator which represents the actual stimulus situation and another which incorporates expectations about it. In contrast to the existing mechanisms, two such indicators would not necessarily vary with the same periodicity as each other or as the stimulus sequence. As a result, the summation of their effects might produce reversals in performance as seen here. Indeed, there is some evidence in the Period x Probability interaction for response time data, described above, that observers do use their expectations in the determination of performance, and that, in this case, expectations did not change at the same rate as stimulus probability.

5. Summary and Conclusions

In summary, it appears that there are problems involved in any attempt to account for the findings of the present experiment in terms of any of the processes. The changes in accuracy and response time between probability conditions are consistent with the ideal observer hypothesis or the target confidence mechanism applied to any of the models. However, only the ideal observer hypothesis applied to Signal Detection Theory can also account for the changes in confidence between blocks although the response stabilisation hypothesis applied to the random walk or accumulator models can account for the confidence data on its own. This finding would appear to indicate some shortcomings in the formulations for confidence in the decision processes. In particular, the role of the criterion regulation mechanisms in the determination of confidence appears to be important.

In addition, when the changes in the response measures within probability conditions are considered as well, the findings are no longer consistent with the ideal observer hypothesis based on either recent or cumulative probability. Instead the reversals in the direction of change of the response measures during blocks suggest that there may be two measures of the stimulus situation contributing to the determination of the criterion, and that these two measures vary at different rates. This general form of mechanism is embodied in the response stabilisation and adaptation-level hypotheses. However, in these two mechanisms, the two indicators of the stimulus sequence are necessarily dependent in this experimental situation, so both indicators will always vary in the same direction and their combination will not give rise to reversals. As a consequence, according to these mechanisms, the initial change in the criterion in response to a step change in probability should continue throughout the block rather than change direction as found in the present data.

It appears that no single process incorporates all of the features necessary to explain these findings. In particular, one aspect which needs some clarification is the formulation for confidence. From the experiments performed so far it appears that the manipulation of a priori probability may not be the most effective way of inducing changes in confidence ratings. In addition, criterion regulation mechanisms such as the adaptation-level hypothesis are relatively insensitive to variations in a priori probability. One possible alternative appears to be changes in discriminability since, on a trial to trial level, confidence is more directly determined by the level of discriminability than by a priori probability or other such factors.

CHAPTER 6

A. INTRODUCTION

In Chapter 5 it was noted that, while some of the mechanisms for criterion regulation, such as the response stabilisation hypothesis, were sensitive to changes in a priori probability, other mechanisms, such as the adaptation-level hypothesis, were relatively unresponsive. In contrast, it was observed that the adaptation-level hypothesis is sensitive to variations in discriminability, with the result that criterion values may vary from trial to trial, but that the response stabilisation hypothesis does not provide any predictions in this situation. Similarly, on a trial to trial basis, some of the mechanisms which may determine confidence are more directly determined by discriminability than by a priori probability.

The effects of variations in discriminability or stimulus intensity have often been examined in studies of signal detection (e.g. Grice & Hunter, 1964; Grice, 1968; Kohfeld, 1968; Gescheider et al., 1968, 1969; Murray, 1970; Lappin & Disch, 1972b). However, only one discrimination experiment is known in which criterion values were examined as discriminability was varied. In an experiment in which observers were asked to judge whether circles lay to the right or left of a central marker, Curry, Nagel and Gai (1977) defined discriminability as the distance between each circle and the marker. They found evidence of variations in the values of β adopted by individual observers as this distance was manipulated, but the pattern of change in β was not consistent across observers. Only accuracy measures were taken in this study and no evidence is available concerning the biassing influence of discriminability on response time and confidence.

B. PREDICTIONS OF THE DECISION PROCESSES

1. The adaptation-level hypothesis

(a) Signal Detection Theory

When the adaptation-level hypothesis is applied to Signal Detection Theory, the position of the cutoff will vary as a direct function of the value of the adaptation level which, in turn, is determined by discriminability. For example, discriminability may be varied in a within-subjects design similar to Experiment 5. Accordingly, it will be assumed that there are three possible levels of discriminability (easy, medium and difficult) for both right and left stimuli, and that these levels are varied so that experimental blocks take one of two forms in alternation. In context RD (right difficult) blocks, right stimuli are of either medium or difficult discriminability while left stimuli are of either easy or medium discriminability. In context RE (right easy) blocks, right stimuli are of either easy or medium discriminability and left stimuli are of either medium or difficult discriminability. The four levels of discriminability used in each type of block occur with equal probability in the block. Prior to the experimental session there are 400 practice trials in which the six levels of discriminability are equiprobable. The experimental session consists of two practice blocks followed by four test blocks each of 120 trials, forming a sequence of contexts RD : RE : RD : RE : RD : RE. Since a change from context RD to context RE for right stimuli necessarily corresponds to the opposite change for left stimuli, predictions for responses to left stimuli vary in the complementary way to those for right stimuli and detailed predictions are presented in terms of right stimuli only. Similarly, since the variation in the response measures is the same at each level of discriminability, only the medium level of discriminability is considered. In addition, for all response measures the variation which results from a change from context RD to context RE is the converse of that from a change from context RE to context RD, so only the change from context RE to context RD is described in detail.

In order to avoid the problem of negative intensities it will be assumed that the distributions of the sensory effect of easy, medium and difficult left stimuli and easy, medium and difficult right stimuli are normal with means of 6, 8, 10, 16, 14 and 12, respectively, and that $\sigma = 1.0$ for all distributions. In this experimental situation, the adaptation level, calculated every 20 trials (i.e. $x = 20$, see Chapter 5) varies as shown in figure 49(a). The adaptation level shows a ramp decrease in context RD. In addition, it can be seen that the magnitude of the variation in the adaptation level decreases with successive blocks. Since the changes in all three response measures are a direct result of the variation in the adaptation level, the changes in the response measures will also decrease in size with successive blocks.

The ramp decrease in the adaptation level in context RD corresponds to a gradual movement of the cutoff from x_{c_1} towards x_{c_2} in figure 41. As the cutoff moves, the probability that observations will favour right responses will increase, so that more right responses will be made, with a consequent increase in the proportion of correct responses to right stimuli. Figure 49(b) shows the variation in the accuracy of the medium discriminability right responses which corresponds to the expected values of the cutoff shown in figure 49(a). There is a ramp increase in accuracy in context RD.

Since response time is an inverse function of the distance between an observation and the cutoff, the changes in the position of the cutoff, as determined by the adaptation-level hypothesis, will give rise to changes in response time. As described in Chapter 5, the average distance between an observation and the cutoff can be estimated by subtracting the cutoff from

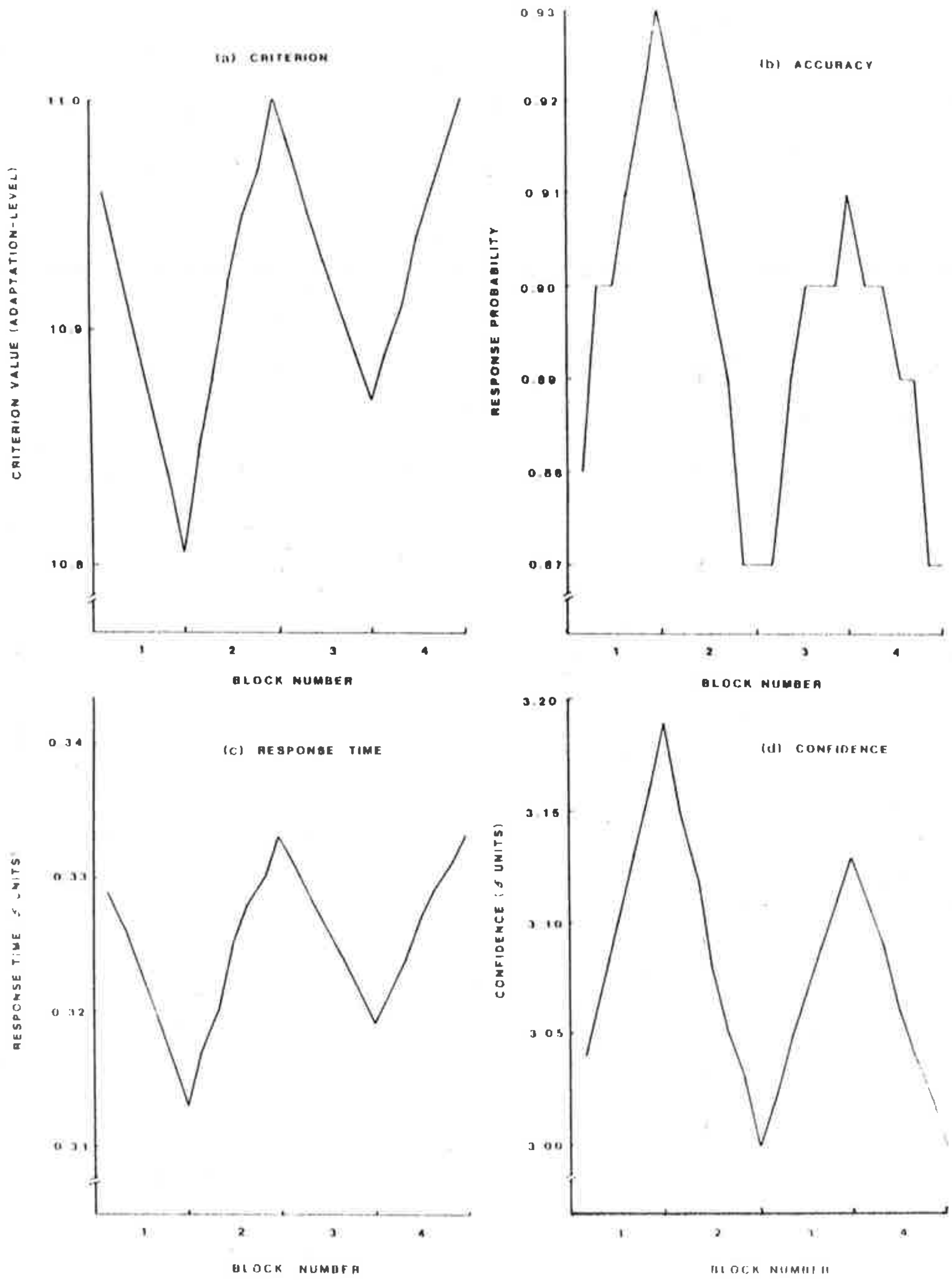


FIGURE 49. Predictions of the adaptation-level hypothesis applied to Signal Detection Theory for right, medium discriminability stimuli when the context of the stimulus takes the values RD : RE : RD : RE in consecutive blocks of 120 trials numbered 1, 2, 3 and 4, respectively. It is assumed that there were 400 practice trials in which the six levels of discriminability were equiprobable. In addition, it is assumed that the distributions of sensory effect of easy, medium and difficult left stimuli and easy, medium and difficult right stimuli are normal with means of 6, 8, 10, 16, 14 and 12, respectively, and that $\sigma=1.0$ for all distributions. The figures show the expected variation in (a) the criterion, β, (b) response probability, (c) response time and (d) confidence.

the mean of the truncated normal distribution corresponding to correct right responses. Figure 49(c) shows the reciprocal of the values of this difference for each cutoff value in figure 49(a). Times for right responses show a ramp decrease in context RD.

Confidence, which is a direct function of the distance between an observation and the cutoff, will vary as shown in figure 49(d). Confidence in right responses shows a ramp increase in context RD.

(b) The random walk model

When the adaptation-level hypothesis is applied to the random walk model in this experimental situation, it may be assumed that the variation in the adaptation level due to the changes in discriminability will give rise to a change in the rate at which evidence in favour of one response or the other is collected. For example, it may be assumed that the distributions of the sensory effect of medium discriminability left stimuli and right stimuli are normal as represented in figure 41. In this case, the decrease in the adaptation level in context RD will correspond to a movement of the indifference point, which marks the change in observations from favouring right to favouring left responses, from x_{c_1} towards x_{c_2} in figure 41. As the value of the indifference point changes, there will be an increase in the number of observations which favour right responses and each observation will tend to yield a large amount of evidence in favour of a right response. In terms of the random walk model, in context RD, the size of the steps (or the drift, Θ) which favour right responses will tend to be large. As a result, the boundary for right responses is likely to be reached first, so more right responses will be made with a consequent increase in accuracy.

Similarly, since the number and magnitude of observations which favour right responses will increase as the adaptation level decreases in context RD, there will be a corresponding decrease in the number of observations required before the boundary for right responses is reached. Therefore, times for right responses will decrease.

As seen in Chapter 2, Ascher (1974) suggested that, in the random walk model, confidence varies as a function of the value of the starting point, the boundaries and the drift (see equation (7)). Since the starting point and the boundaries do not change in this situation, confidence is directly related to the drift, Θ. Therefore, when the adaptation level decreases in context RD, producing an increase in Θ, confidence will increase.

(c) The accumulator model

As in the random walk model, when the adaptation-level hypothesis is applied to the accumulator model in this experimental situation, it can be assumed that changes in the adaptation level give rise to variations in the rate at which evidence in favour of one response or the other will be collected. For example, the decreasing adaptation level in context RD will result in an increase in the number and magnitude of observations which favour a right response. Therefore, the qualitative predictions of this process are identical to those described above for the random walk model.

2. The target confidence mechanism

In Chapter 2 it was noted that it is not possible to apply the target confidence mechanism to Signal Detection Theory, so only the variable sample models will be considered.

(a) The accumulator model

When the target confidence mechanism is applied to the accumulator model it determines the relative values of the criteria in the primary decision process. For example, when right stimuli are of either medium or difficult discriminability in context RD, the magnitude of observations which favour right responses will tend to be small, so that many observations will be needed to satisfy the criterion for right responses. In this case, it is likely that some observations which favour left responses will also be collected before the right response is made, resulting in lower confidence in the right response. As a consequence, confidence in right responses will tend to be less than target confidence and, on the majority of trials, underconfidence will be stored in the appropriate control accumulator. Therefore, it is likely that the criterion for underconfidence will be satisfied before the criterion for overconfidence, so the criterion for right responses in the primary decision process will increase.

The increase in the relative value of the criterion for right responses in context RD will result in fewer right responses being made, so accuracy will be low. At the same time, a large number of observations will be needed to satisfy the high criterion, so times for right responses will tend to be long. Confidence will also vary with the relative values of the criteria. For example, in context RD, when the criterion for right responses is high, the maximum possible confidence in right responses will be equal to the high criterion value. Therefore, confidence in right responses will tend to be high.

(b) The random walk model

As seen in Chapter 2, in order to be able to apply the target confidence mechanism to the random walk model it is necessary to assume that the value of the desired likelihood ratio diminishes over time. This corresponds to a decrease in the distance between the boundaries with each successive observation taken before a response is made (see figure 8). In this form of the model, the target confidence mechanism determines the relative distance between each boundary and the starting point. In the present experimental situation, variations in discriminability may be interpreted as giving rise to changes in the relative caution with which the alternative responses are made. Therefore, in accordance with Relative Judgment Theory it may be assumed that the boundaries are adjusted relative to the constant starting point (Ascher, 1974; Link & Heath, 1975; Link, 1978a). Since the levels of caution for alternative responses may vary in opposite ways, it may be assumed that each boundary is adjusted independently (Link & Heath, 1975). For example, when right responses are of either medium or difficult discriminability in context RD, observations which favour right responses will tend to yield only small amounts of evidence, so the drift or step size of the random walk will be small. Therefore, right responses will be made only after many observations and, since the boundary for right responses diminishes with each observation, confidence will tend to be less than target confidence. As a consequence, the boundary for right responses will be likely to move away from the starting point. This process is analogous to the target confidence mechanism applied to the accumulator model, and its predictions are qualitatively identical to those for the accumulator model as described above.

3. Summary

The qualitative predictions of the adaptation-level hypothesis and the target confidence mechanism applied to each of the decision models are summarised in figure 50. All of the predictions are given in terms of responses

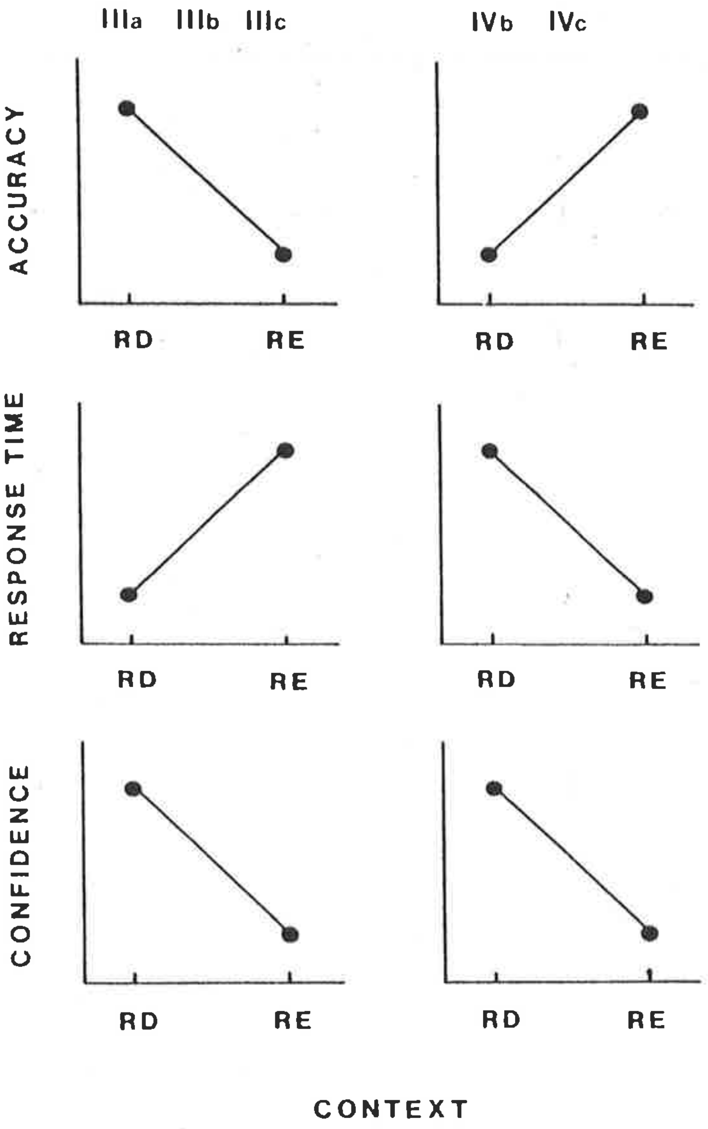


FIGURE 50. A summary of the qualitative predictions of the adaptation-level hypothesis and the target confidence mechanism applied to each of the decision models. All of the predictions are given in terms of responses to right stimuli of medium discriminability. Context RD indicates a block in which right stimuli are of either medium or difficult discriminability while left stimuli are of either easy or medium discriminability. Context RE indicates a block in which right stimuli are of either easy or medium discriminability while left stimuli are of either medium or difficult discriminability. The decision processes are labelled as in figure 48(b).

to right stimuli of medium discriminability. Context RD indicates a block in which right stimuli are of either medium or difficult discriminability while left stimuli are of either easy or medium discriminability. Context RE indicates a block in which right stimuli are of either easy or medium discriminability and left stimuli are of either medium or difficult discriminability.

From figure 50 it appears that the processes can be grouped on the basis of the mechanism for criterion regulation. The accuracy and response time predictions of the adaptation-level hypothesis applied to any of the decision models are opposite to those of the target confidence mechanism applied to the random walk or accumulator models. On the other hand, both of the mechanisms for criterion regulation provide the same predictions about confidence. Therefore, the processes may be compared on the basis of their accuracy and response time predictions in relation to the confidence prediction. For example, the adaptation-level hypothesis predicts an inverse relationship between response time and confidence between contexts, while the target confidence mechanism predicts a direct relationship in this situation.

C. EXPERIMENT 6

1. Introduction

In order to evaluate these predictions, measures of accuracy, response time and confidence were taken when bias was manipulated by variations in discriminability according to a within-subjects design.

2. Method

(a) Stimuli

The stimuli were white pi-shaped figures identical to those used in Experiment 3 and described in Chapter 4, section A.2(a) except that three levels of discriminability were used. The longer line was always 187 mm in length and the shorter line was either 186.5 mm, 185.5 mm or 184.5 mm in length. These levels of discriminability were labelled 1 (difficult), 3 (medium) and 5 (easy) respectively, to correspond to previous experiments in the thesis.

(b) Apparatus

The apparatus was identical to that used in Experiment 3 and described in Chapter 4, section A.2(b).

(c) Observers

The 20 observers, 10 men and 10 women, were students enrolled in the first year psychology course at the University of Adelaide, whose participation was credited towards a course requirement. They ranged in age from 17 to 20 years and all declared themselves to be right-handed. All observers were naive with respect to the aims of the experiment.

(d) Design

Two groups of 10 observers were formed, each comprising five men and five women. The groups performed under different order conditions with each observer being tested individually for approximately one hour.

In both order conditions stimuli were presented in 10 blocks: 3 practice blocks each of 60 trials preceded 7 experimental blocks each of 120 trials. For both groups of observers, in each block of trials whether practice or experimental, the occurrence of the longer line on the right or the left was equiprobable ($p(R) = 0.5$) and right and left stimuli appeared in a random sequence. In addition, within all blocks of practice trials and the first block of experimental trials the three levels of discriminability occurred on the right and the left with equal probability. The remaining 6 blocks of experimental trials took one of two forms:

RD: Discriminability levels 1 (difficult) and 3 (medium) occurred when the right line was longer and discriminability levels 3 (medium) and 5 (easy) occurred when the left line was longer.

RE: Discriminability levels 3 (medium) and 5 (easy) occurred when the right line was longer and discriminability levels 1 (difficult) and 3 (medium) occurred when the left line was longer.

One group of observers, Order RD/RE, performed the last six experimental blocks in the order RD : RE : RD : RE : RD : RE, while the other group of observers, Order RE/RD, performed the last six experimental blocks in the order RE : RD : RE : RD : RE : RD. The two appropriate levels of discriminability occurred equally often on the side to which they were assigned in any given block of trials and observers were given no indication that the context was being manipulated. The six experimental blocks were performed in an uninterrupted sequence. For both groups of observers the first two blocks were treated as practice and only the last four experimental blocks were considered in the analyses.

(e) Procedure

The procedure was identical to that used in Experiment 4 and described in Chapter 4, section A.2(e).

3. Results

Analyses of variance with Order (RD/RE or RE/RD, as defined in section C.2(d), Period (first two experimental blocks or second two experimental blocks), Context (block type RD or type RE, as defined in section C.2(d)), Half (first or second half of a block) and Response (right or left) as factors were used to analyse data from responses to the common level of discriminability, level 3 (medium) from the last four experimental blocks. Analyses of variance were carried out according to a crossed factorial design with repeated measures on Period, Context, Half and Response. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses.

(a) Accuracy

The analysis of accuracy data showed significant interactions of Period x Order ($F(1,18) = 4.73$, MSerror = 0.09, $p < .05$) and Half x Response x Order ($F(1,18) = 14.13$, MSerror = 0.03, $p < .01$).

From table 19 it can be seen that, in Order RD/RE there was a nonsignificant decrease in accuracy from Period 1 to Period 2. In contrast, in Order RE/RD there was a nonsignificant increase in accuracy from Period 1 to Period 2. This pattern gave rise to the Period x Order interaction.

	PERIOD	
ORDER	1	2
RD/RE	98.67	97.32
RE/RD	97.25	97.74

TABLE 19. Percentage of stimuli for which responses were correct in Periods 1 and 2 of Order RD/RE and Order RE/RD. The percentages were taken over the 10 observers in each order.

The Half x Response x Order interaction is shown in figure 51. In Order RD/RE, the accuracy of right responses decreased while the accuracy of left responses increased from the first to the second half of blocks. Conversely, in Order RE/RD the accuracy of right responses increased while the accuracy of left responses decreased from the first to the second half of blocks.

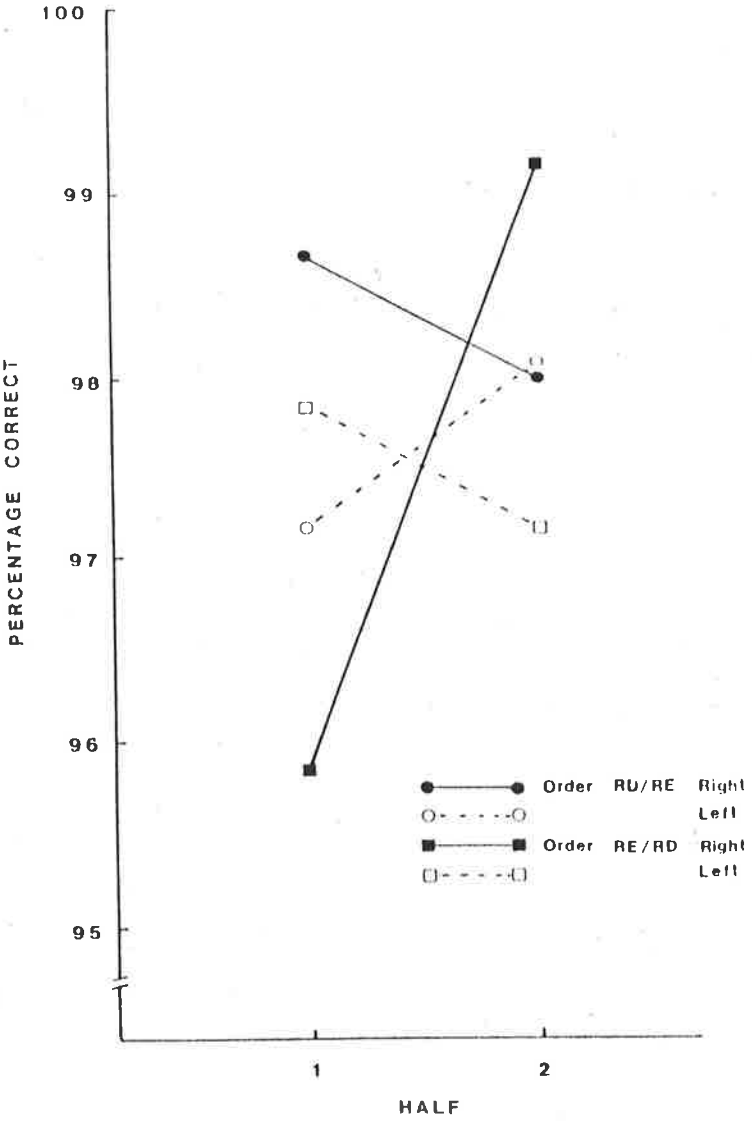


FIGURE 51. Percentage of right stimuli (solid line) and left stimuli (broken line) for which responses were correct in Half 1 and Half 2 of blocks in Order RD/RE (circles) and Order RE/RD (squares). The percentages were taken over the 10 observers in each order.

(b) Response time

The analysis of response time data showed a significant main effect of Response ($F(1,18) = 10.37$, $MSerror = 13599$, $p < .01$).

The main effect of Response was significant because right responses were made more quickly than left responses. The mean times for right and left responses were 787.40 msec and 829.39 msec, respectively.

(c) Confidence

The analysis of confidence data showed significant interactions of Half x Response ($F(1,18) = 6.07$, $MSerror = 0.04$, $p < .05$), Period x Half x Response ($F(1,18) = 4.98$, $MSerror = 0.03$, $p < .05$) and Period x Context x Half x Response x Order ($F(1,18) = 6.28$, $MSerror = 0.05$, $p < .05$).

From table 20 it can be seen that confidence in right responses decreased while confidence in left responses increased from the first to the second half of blocks. This gave rise to the Half x Response interaction.

RESPONSE	HALF 1	2
RIGHT	3.14	3.19
LEFT	3.03	2.97

TABLE 20. Mean confidence in correct right and left responses in Half 1 and Half 2 of blocks. The means were taken over 20 observers, the 10 observers in Order RD/RE and the 10 observers in Order RE/RD.

The Period x Half x Response interaction is shown in figure 52 where it can be seen that, although none of the changes in confidence from the first to the second half of blocks was significant, right responses in the second half of blocks in Period 2 were significantly more confident than left responses in the second half of blocks in Periods 1 and 2.

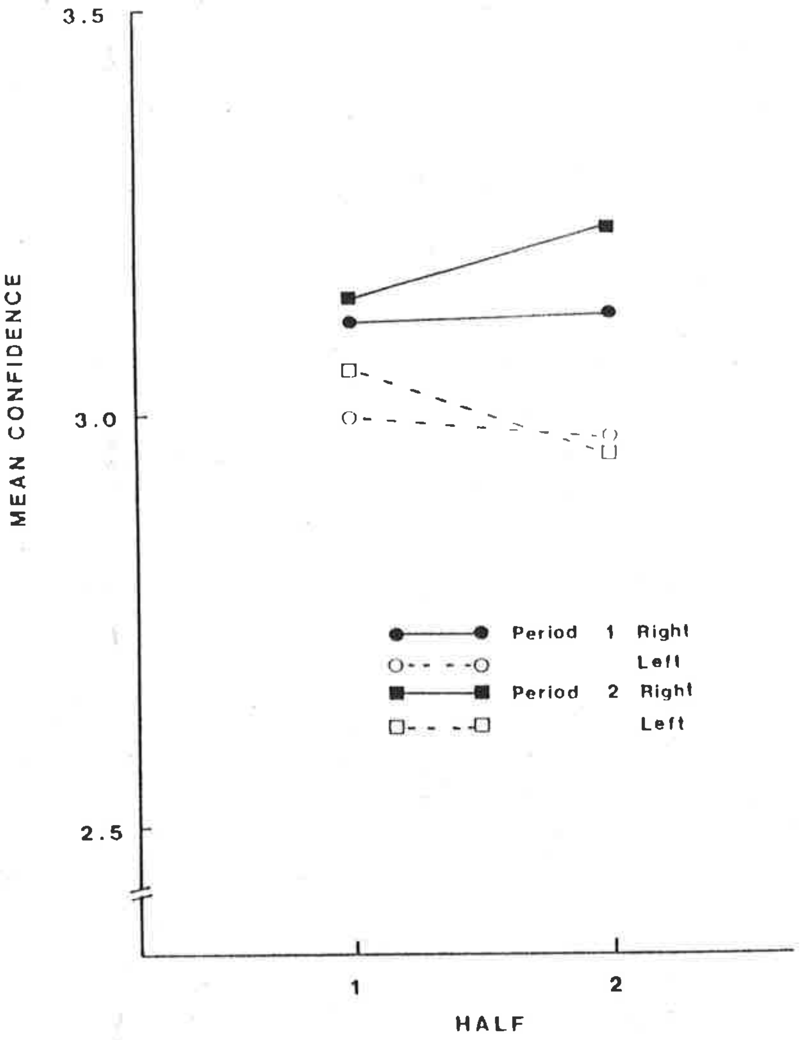


FIGURE 52. Mean confidence in correct right responses (solid line) and left responses (broken line) in Half 1 and Half 2 of blocks in Period 1 (circles) and Period 2 (squares). The means were taken over 20 observers, the 10 observers in Order RD/RE and the 10 observers in Order RE/RD.

The Period x Context x Half x Response x Order interaction is shown in table 21 where it is apparent that confidence in right and left responses did not vary systematically between contexts. In addition, there was no consistent change in confidence as the session progressed.

CONTEXT		RD				RE			
HALF		1		2		1		2	
RESPONSE		R	L	R	L	R	L	R	L
	PERIOD								
ORDER RD/RE	1	3.10	3.09	3.08	2.99	3.19	2.81	3.13	2.84
	2	3.15	2.99	3.22	2.92	3.16	2.98	3.33	2.93
ORDER RE/RD	1	3.14	3.02	3.13	3.18	3.06	3.09	3.16	2.91
	2	3.03	3.18	3.17	2.97	3.27	3.09	3.22	3.00

TABLE 21. Mean confidence in correct right and left responses in the first and second halves of blocks of Context RD and of Context RE in each period for Order RD/RE and for Order RE/RD. The means were taken over the 10 observers in each order.

(d) Confidence and response time

For each observer, confidence in all responses, (both correct and incorrect), was plotted against the corresponding response time values and straight lines were fitted. Separate plots were made for right and left responses in each block. Of the total of 160 plots, 114 of the slopes were negative, indicating a predominantly inverse relationship between confidence and response time within blocks ($p < .001$, binomial test). The inverse relationship was still evident when the 40 plots for right and left responses in each context were considered. In Context RD, the numbers of negative slopes for right and left responses were 29 and 27 respectively ($p < .01$, $p < .05$, respectively, binomial test). In Context RE, the numbers of negative slopes for right and left responses were 26 and 32 respectively ($p < .05$, $p < .001$, respectively, binomial test). Therefore, there is an inverse relationship between confidence and response time for right and left responses within blocks, irrespective of context.

4. Discussion

It appears that the effects which reached significance in the analysis of accuracy, response time and confidence did not result from the manipulation of discriminability. For example, both the Period x Order interaction and the Half x Response x Order interaction which were found in accuracy measures may be explained in terms of differences in the performance of individual observers in the two order conditions. Therefore, these findings do not correspond to explicit predictions of any of the processes.

Similarly, the Response main effect shown by response time data provides further evidence for the pre-existing preference of right-handed observers for right responses which were made more quickly. This finding can be accommodated by any of the processes but it is not specifically predicted by any of them.

In addition, confidence data showed several significant interactions which did not appear to be a simple result of the variations in discriminability. It appears that the interactions of Half x Response, Half x Response x Order and Period x Context x Half x Response x Order arose as a result of unsystematic variations in responding which occurred as the experimental session progressed. Therefore, although these findings are not inconsistent with the decision processes, they are not predicted.

From these findings it seems that the manipulation of discriminability used in this experiment did not change the relative caution of responding. This may indicate that this form of manipulation does not have a strong influence on performance. However, a more likely explanation is suggested by the accuracy data. Accuracy rates for both right and left responses at the common level of discriminability were consistently high (> 95%) indicating that this was not a very difficult discrimination. Consequently, these findings may merely show the effects of a ceiling in performance which was reached at easy levels of discriminability. If this were the case, these data are consistent with the results obtained in Experiments 3 and 4 where the three response measures, accuracy, time and confidence, tended to an asymptote at discriminability levels 3, 4 and 5. Therefore, it appears that no conclusions about the effectiveness of this manipulation of discriminability can be drawn from these findings.

D. EXPERIMENT 7

1. Introduction

The experiment was repeated but, in this case, a smaller stimulus difference was used for the medium level of discriminability common to right and left stimuli. Since there was no evidence of asymptotic performance at Discriminability level 2 in Experiments 3 and 4, this was chosen as the common, medium level. As there were no significant main effects due to the order in which observers performed the two types of context in Experiment 6, only one order condition was considered in this experiment.

2. Method

(a) Stimuli

The stimuli were white pi-shaped figures identical to those used in Experiment 3 and described in Chapter 4, section A.2(a) except that three levels of discriminability were used. The longer line was always 187 mm in length and the shorter line was either 186.5 mm, 186.0 mm or 184.5 mm in length. These levels of discriminability were labelled 1 (difficult), 2 (medium) and 5 (easy), respectively, to correspond to previous experiments in the thesis.

(b) Apparatus

The apparatus was identical to that used in Experiment 3 and described in Chapter 4, section A.2(b).

(c) Observers

The 10 observers were women enrolled in the first year psychology course at the University of Adelaide, whose participation was credited towards a course requirement. They ranged in age from 17 to 30 years and all declared themselves to be right-handed. All observers were naive with respect to the aims of the experiment.

(d) Design

The design was identical to that of Experiment 6 described in section C.2(d) except that all observers performed the experimental blocks in the order RD : RE : RD : RE : RD : RE.

(e) Procedure

The procedure was identical to that used in Experiment 3 and described in Chapter 4, section A.2(e).

3. Results

Analyses of variance with Period (first two experimental blocks or second two experimental blocks), Context (block type RD or type RE, as defined in section C.2(d)), Half (first or second half of a block) and Response (right or left) as factors were used to analyse data from the responses to the common level of discriminability, level 2 (medium), from the last four experimental blocks. Analyses of variance were carried out according to a crossed factorial design with repeated measures on all factors. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses.

(a) Accuracy

The analysis of accuracy data showed a significant interaction of Period x Half (F(1,9) = 8.64, MSerror = 0.04, $p < .05$).

From table 22 it can be seen that, in Period 1, accuracy increased from the first to the second half of blocks. In contrast, in Period 2, accuracy decreased from the first to the second half of blocks.

	PERIOD	
HALF	1	2
1	95.83	96.83
2	97.66	95.16

TABLE 22. Percentage of stimuli for which responses were correct in Half 1 and Half 2 of blocks in Periods 1 and 2. The percentages were taken over the 10 observers.

(b) Response time

The analysis of response time data showed a significant Context x Half x Response interaction (F(1,9) = 12.29, MSerror = 3545, p < .01).

The Context x Half x Response interaction is shown in figure 53. It is apparent that response time varied both between and within contexts.

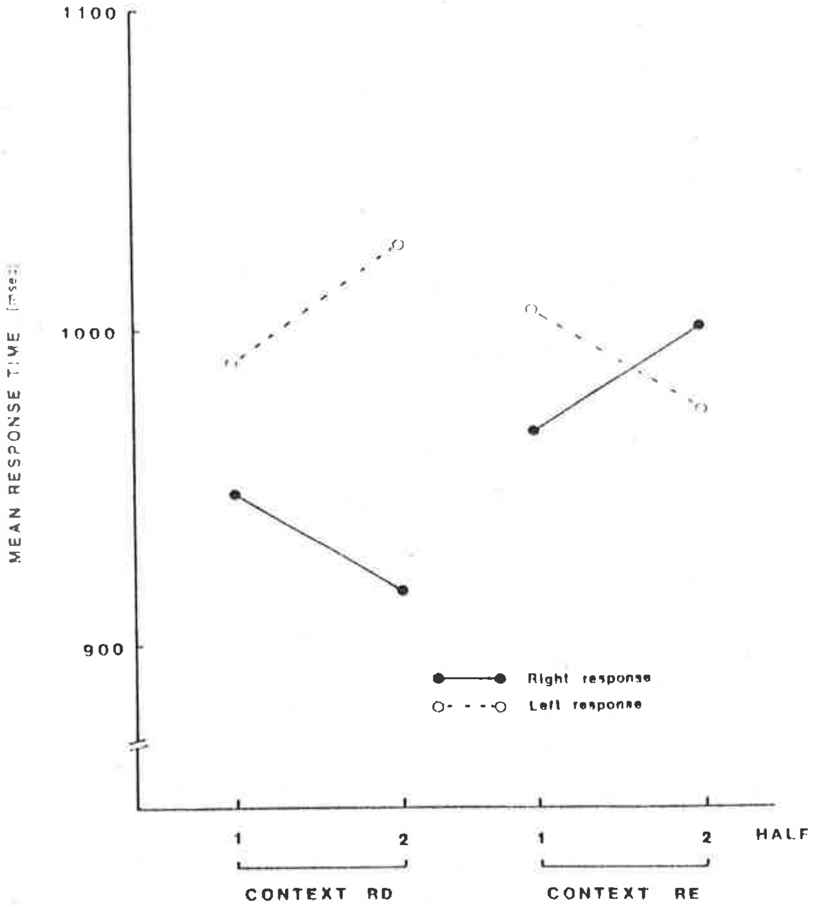


FIGURE 53. Mean times for correct right responses (solid line) and left responses (broken line) in Half 1 and Half 2 of blocks of Context RD and of Context RE. The means were taken over the 10 observers.

When discriminability changed from the second half of Context RD to the first half of Context RE, there was a significant increase in times for right responses and a significant decrease in times for left responses. Conversely, when discriminability changed from the second half of Context RE to the first half of Context RD, there was a significant decrease in times for right responses and a significant increase in times for left responses. At the same time, times for right responses showed a significant decrease and times for left responses showed a significant increase from Half 1 to Half 2 of blocks of Context RD. Conversely, times for right responses increased significantly and times for left responses decreased significantly from Half 1 to Half 2 of blocks of Context RE.

(c) Confidence

The analysis of confidence data showed a significant Context x Half x Response interaction ($F(1,9) = 10.89$, $MSerror = 0.02$, $p < .01$).

The Context x Half x Response interaction is shown in figure 54 where it is apparent that confidence changed both between and within contexts. When discriminability changed from the second half of Context RD to the first half of Context RE there was a nonsignificant increase in confidence for right responses and a significant decrease in confidence for left responses. When discriminability changed from the second half of Context RE to the first half of Context RD there was a nonsignificant decrease in confidence for right responses and a significant increase in confidence for left responses. On the other hand, confidence in right responses showed a nonsignificant increase and confidence in left responses showed a significant decrease from Half 1 to Half 2 of Context RD. In Context RE, confidence in right responses showed a nonsignificant decrease while confidence in left responses did not change from Half 1 to Half 2.

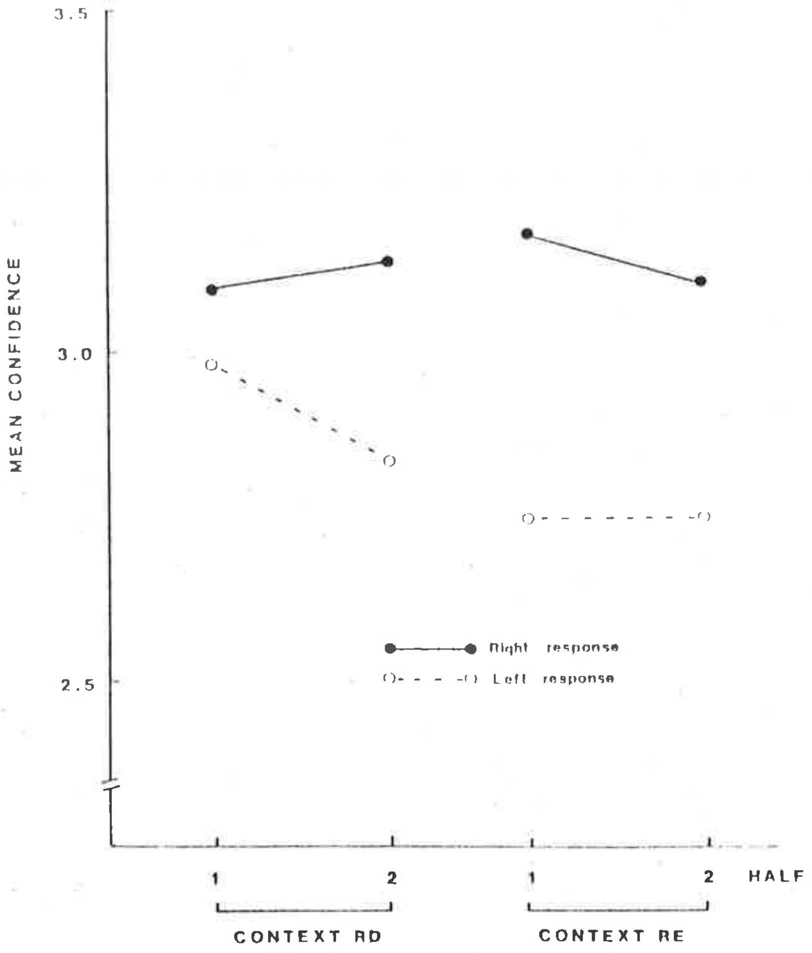


FIGURE 54. Mean confidence in correct right responses (solid line) and left responses (broken line) in Half 1 and Half 2 of blocks of Context RD and of Context RE. The means were taken over the 10 observers.

(d) Confidence and response time

For each observer, confidence in all responses, (both correct and incorrect), was plotted against the corresponding response time values and straight lines were fitted. Separate plots were made for right and left responses in each block. It was found that 66 of the 80 best fitting straight lines had negative slopes ($p < .001$, binomial test), indicating a predominantly inverse relationship between confidence and response time within blocks. The inverse relationship was still evident when the 20 plots for right and left responses in each context were considered separately. In Context RD, the numbers of negative slopes for right and left responses were 20 and 15 respectively ($p < .001$, $p < .05$, respectively, binomial test). In Context RE, the numbers of negative slopes for right and left responses were 14 and 17 respectively ($p > .05$, $p < .01$, respectively, binomial test). Therefore, for all except right responses in Context RE, there was a significant inverse relationship between confidence and response time for right and left responses within blocks, irrespective of context.

4. Discussion

The Period x Half interaction in accuracy data does not appear to result from the manipulation of discriminability. Although, according to the adaptation-level hypothesis, the size of the variation in response measures should decrease with successive changes in context, this form of consistent change across blocks would be expected to produce a Period x Context x Half interaction and does not appear to have given rise to the 2-way interaction. Instead, the Period x Half interaction may be interpreted in terms of fluctuations in arousal which are not related to the organisation of the experiment into contexts or to the nature of the stimulus material. Since these changes were not found in either response time or confidence data and the greatest variation in accuracy was less than 2%, this interaction reflects only minor changes in the efficiency of performance.

The effects of the manipulation of discriminability are evident in the Context x Half x Response interaction which was significant for response time and confidence. Therefore, in contrast to the findings of Curry et al. (1977), in this experiment there is evidence of variations in criterion values which were consistent across observers.

From a comparison of the changes in response time between contexts with the predictions of the decision processes, it is apparent that the increase in times for right responses and the decrease in times for left responses from Context RD to Context RE is consistent with the adaptation-level hypothesis but not with the target confidence mechanism.

The pattern of confidence data was less clear. The lack of any change in confidence in right responses from Context RD to Context RE contrasts with the predictions of all of the decision processes according to which confidence should vary between contexts. Similarly, the decrease in confidence in left responses from Context RD to Context RE is opposite to their predictions. While this finding was significant for only one response and therefore must be interpreted with some reservation, it would appear to suggest that, independent of the way in which the criterion is regulated, observers may modify their internally experienced level of confidence before reporting it. Intuitively, it seems likely that observers will judge their confidence in the medium level of discriminability of a stimulus relative to their confidence in the more extreme level of discriminability of the same stimulus. Therefore, in Context RD, medium discriminability left stimuli would be judged in comparison to easy discriminability left stimuli. Since it would be expected that responses to easy discriminability stimuli will be highly confident, if the observer uses this high level of confidence as an anchor and contrast effects are operating (Helson, 1964), he will adjust his internally experienced confidence in responses to medium discriminability stimuli downwards. Conversely, in Context RE it would be expected that observers will adjust their confidence in responses to medium discriminability stimuli upwards. This would give rise to a pattern of change in confidence in agreement with these data but opposite to the predictions of the decision processes.

Alternatively, it is possible that, in the case where the stimulus is paired with a difficult discrimination which will not always be accurately discriminated, the observer assesses this stimulus as less probable. Therefore, an adaptation-level type of mechanism may give rise to an effective increase in the discriminability of this stimulus and hence, confidence in the corresponding response will be higher.

Similarly, there are problems involved in the explanation of the different patterns of change in response time and confidence within contexts. The adaptation-level hypothesis predicts that changes within contexts will take the form of a continuation of the changes between contexts. In contrast, the target confidence mechanism does not yield any specific predictions about these changes. Instead, the nature of the changes within conditions depends on the state of the decision process at the time of each change in context. For example, if there is damping in the decision process, the effect of the change in context may not be apparent until some time after it occurs, so the change between contexts would continue and then reverse at some point later in the block. In this case, the change within contexts would take the form of a continuation of the change between contexts. Alternatively, the change in performance due to a change in context may be immediate so that the change within contexts would be opposite to the change between contexts. As a consequence, according to this process the changes in response measures within contexts may vary from one block to the next.

For response time, the changes within contexts form a continuation of the changes between contexts as predicted by the adaptation-level hypothesis. In contrast, the change in confidence within contexts is in the opposite direction to the change between contexts. This is in direct contrast to the adaptation-level hypothesis but it is not inconsistent with the target confidence mechanism.

5. Summary and Conclusions

It appears that there are problems involved in any attempt to account for the changes in response measures both within and between contexts. When only the changes between contexts are considered, response time data are consistent with the adaptation-level hypothesis applied to any of the decision models, while confidence data do not correspond to the predictions of any of the processes. Similarly, the changes in response time within contexts are predicted by the adaptation-level hypothesis. On the other hand, the changes in confidence within contexts are in direct contrast to those predicted by the adaptation-level hypothesis but they are not inconsistent with the target confidence mechanism.

Despite the equivocal nature of these findings, an examination of the changes between contexts alone appears to have some implications for the nature of formulations for confidence. In this study, confidence varies directly with response time between conditions, while in Experiment 5 there was an inverse relationship between confidence and response time between conditions. It is apparent that any formulation for confidence must be able to account for these different relationships between confidence and response time.

Although these data provide some evidence towards an evaluation of the processes, any conclusions must be tentative. Since accuracy measures did not show any effect due to the manipulation of discriminability, it appears that this form of manipulation does not exert a strong influence on performance. Indeed, it seems that manipulations in which observers are not specifically instructed about the nature of the changes often produce confusing results. This has been the case in the experiments in this thesis in which a priori probability was varied as well as in these studies. A more explicit manipulation may be required to produce consistent effects on performance, perhaps by the use of instructions rather than changes in stimulus properties.

CHAPTER 7

A. INTRODUCTION

To this point in the thesis, the predictions of the decision processes have been compared with the findings of experiments in which bias or relative caution was manipulated by varying either a priori probability or discriminability. In these experiments observers were not informed of the changes which would occur in these stimulus properties. As noted in Chapter 2, another manipulation which might be expected to have equivalent effects on performance is the variation of instructions. The extent to which each mechanism for criterion regulation predicts the exact change in performance to result from any of these manipulations depends on the precise form of the mechanism. In particular, the ideal observer hypothesis predicts that all of these manipulations will give rise to equivalent changes, while in the other three mechanisms, the contribution of stimulus probabilities and discriminability is quite explicit while the role of instructions is not as well defined. At the same time, no discrimination studies are known in which the effects of the different manipulations have been compared, and there is little empirical evidence on the effects of instructions in discrimination tasks.

Although the use of instructions to produce changes in relative caution has been a focus of attention in the study of three-category tasks (e.g. Brown, 1910; Fernberger, 1914a,b, 1931), only one study is known in which instructions were varied in an attempt to vary relative caution in a two-category discrimination task. As described in Chapter 1, Festinger (1943a,b) found that there were systematic changes in the PSE as instructions changed from favouring one response to favouring the alternative. However, details of response time and confidence were not given in this study.

B. PREDICTIONS OF THE DECISION PROCESSES

Neither the response stabilisation hypothesis nor the adaptation-level hypothesis is designed to provide predictions about the effects of instructions, so only the ideal observer hypothesis and the target confidence mechanism will be considered.

In the experimental situation to be considered, there are two sets of instructions. One set (set RA, right accurate) is intended to increase the relative caution with which right responses are made by stressing the accuracy of right responses and the speed of left responses. The other set (set RS, right speed) is intended to reduce the relative caution with which right responses are made by stressing the speed of right responses and the accuracy of left responses. As a consequence, when there is an increase in the relative caution of right responses there will be a decrease in the relative caution of left responses and vice versa. Therefore, predictions for left responses will vary in the complementary way to those for right responses, so detailed predictions are presented in terms of right responses only. In addition, since the two sets of instructions will have complementary effects on performance, only the predictions for set RA are considered.

1. The ideal observer hypothesis

(a) Signal Detection Theory

Although the effects of instructions do not have any direct representation in the ideal observer hypothesis, it seems likely that variations in instructions will produce changes in the subjective costs and payoffs associated with each decision. Therefore, if the observer is trying to maximise the expected value of his decisions, the value of the cutoff will vary according to the subjective costs and payoffs which are set up by the instructions.

For example, when instructions in set RA stress the accuracy of right responses, the subjective value of making a right response will be low compared with the subjective value of making a left response. Therefore, it would be expected that the cutoff will move from position x_{c_2} towards position x_{c_1} in figure 41, so that observations are less likely to fall on the side of the cutoff corresponding to right responses. As a result, fewer right responses will be made, so accuracy, or the percentage of right stimuli for which responses are correct will tend to be low. As seen above, response time is an inverse function of the distance between an observation and the cutoff. Therefore, in set RA, when the cutoff moves towards x_{c_1} (see figure 41), observations which favour right responses will tend to lie close to the cutoff, so times for right responses will be long. At the same time, since confidence varies as a direct function of the distance between an observation and the cutoff, confidence will tend to be low.

(b) The random walk model

In the random walk model, changes in relative caution due to variations in instructions can be conceived of as producing changes in the starting point. For example, in set RA, the starting point may be assumed to move away from the boundary for right responses.

This movement will give rise to changes in accuracy, response time and confidence. In set RA, when the starting point lies at a distance from the boundary for right responses, few right responses will be made, so accuracy will tend to be low. Since response time and confidence vary as a direct function of the distance between the starting point and the boundary, in set RA, right responses will tend to be slow and highly confident (see equation (7)).

(c) The accumulator model

Changes in the subjective costs and payoffs associated with each decision will produce changes in the relative values of the criteria for the two responses in the primary decision process of the accumulator model. For example, in set RA there will be an increase in the relative value of the criterion for right responses. Therefore, the qualitative predictions of this process are identical to those for the random walk model described above.

2. The target confidence mechanism

As seen above, the target confidence mechanism may be applied to either the random walk model or the accumulator model. In both cases, when instructions produce changes in the relative caution of responses, it may be assumed that target confidence for a response will vary as a direct function of the level of caution for that response. For example, in set RA when caution in right responses is high, target confidence for right responses may be assumed to be high. Therefore, since criterion values change in accordance with target confidence, the criterion for right responses will increase. As a result, the qualitative predictions of the target confidence mechanism will be the same as those described above for the ideal observer hypothesis applied to either of these decision processes.

3. Summary

The qualitative predictions of the ideal observer hypothesis and the target confidence mechanism applied to each decision model are summarised in figure 55. In the figure, all of the predictions are given in terms of the influence of instructions on right responses. Instructional set RA indicates that instructions stressed the accuracy of right responses and the speed of left responses, and instructional set RS indicates that instructions stressed the speed of right responses and the accuracy of left responses.

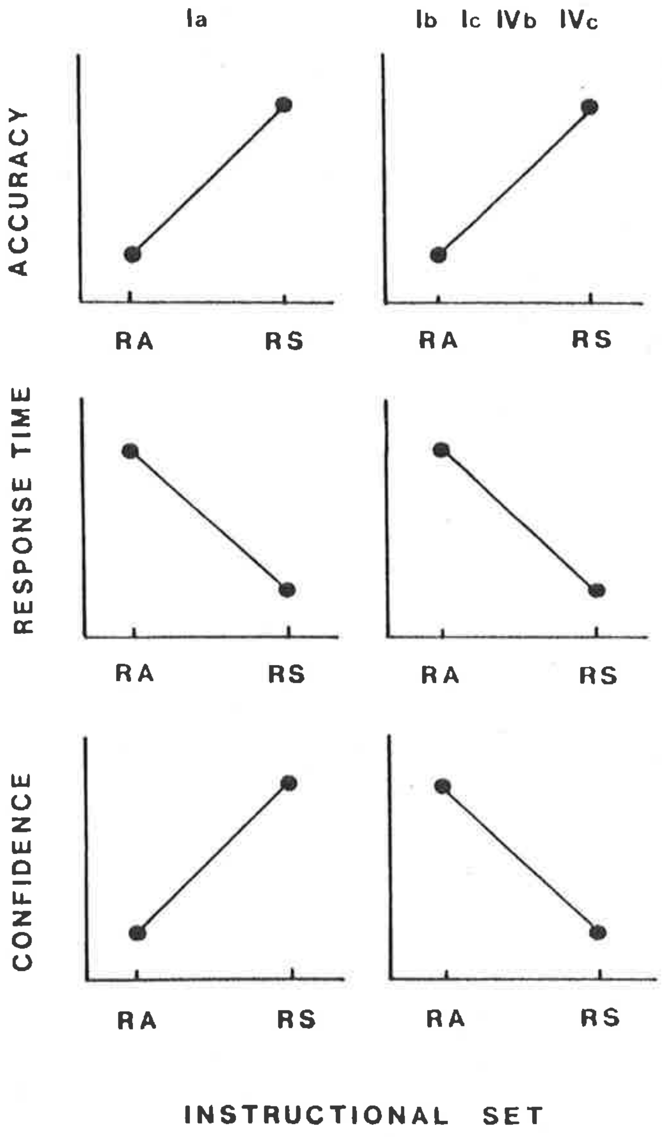


FIGURE 55. A summary of the qualitative predictions of the ideal observer hypothesis and the target confidence mechanism applied to each of the decision models. All of the predictions are given in terms of the influence of instructions on right responses. Instructional set RA indicates that instructions stressed the accuracy of right responses and the speed of left responses, and instructional set RS indicates that instructions stressed the speed of right responses and the accuracy of left responses. The decision processes are labelled as in figure 48(b).

From the figure it is apparent that the decision processes may be differentiated on the basis of confidence predictions. All of the decision processes predict that accuracy will be lower and response times longer under set RA than under set RS. However, when the ideal observer hypothesis is applied to Signal Detection Theory, it is predicted that confidence will be lower under set RA than under set RS. In contrast, when either the ideal observer hypothesis or the target confidence mechanism is applied to the random walk model or the accumulator model it is predicted that confidence will be higher under set RA than under set RS.

C. EXPERIMENT 8

1. Introduction

No studies are known in which all three response measures have been taken when instructions have been varied in order to manipulate relative caution in a two-category discrimination task. Therefore, an experiment was carried out in which measures of accuracy, response time and confidence were taken when the relative caution implied by instructions was varied according to a within-subjects design. Since the confidence measures were of greatest importance in the evaluation of the processes, discriminability was varied within each instructional set in an attempt to increase the range of confidence ratings.

2. Method

(a) Stimuli

The stimuli were white pi-shaped figures identical to those used in Experiment 3 and described in Chapter 4, section A.2(a) with the five levels of discriminability as defined therein.

(b) Apparatus

The apparatus was identical to that used in Experiment 3 and described in Chapter 4, section A.2(b) with one addition. On top of the video monitor, directly above the stimulus display, was a black board on which were positioned a red and a green neon lamp and a buzzer. The lamps were 1 cm in diameter and 15 cm apart, centre to centre. The green lamp, on the right side of the board, was labelled 'right accurate' and the red lamp, on the left side of the board, was labelled 'left accurate'. The buzzer was placed in the centre of the board. The PDP 8/E computer controlled the illumination of either lamp and the sounding of the buzzer.

(c) Observers

The 20 observers were women enrolled in the first year psychology course at the University of Adelaide, whose participation was credited towards a course requirement. They ranged in age from 18 to 21 years and all declared themselves to be right-handed. Observers had accurate red-green colour discrimination and all were naive with respect to the aims of the experiment.

(d) Design

Two groups of ten observers were formed. The groups performed under different order conditions with each observer being tested individually for approximately one hour.

In both order conditions, stimuli were presented in 10 blocks: 4 practice blocks each of 50 trials preceded 6 experimental blocks each of 100 trials. For both groups of observers and within each block of trials, both practice and experimental, the occurrence of the longer line on the right or the left was equiprobable ($p(R) = 0.5$) and the sequence was random. In addition, each of the five levels of discriminability occurred equally often for right stimuli and for left stimuli within each block of trials and the sequence was random. Each block of trials was performed under one of two types of instructional set:

RA. Respond as accurately as possible when the right line is longer and as quickly as possible when the left line is longer.

RS. Respond as quickly as possible when the right line is longer and as accurately as possible when the left line is longer.

One group of observers, Order RA/RS, performed the first practice block under instructional set RA and the second practice block under instructional set RS. Thereafter, instructions alternated between consecutive blocks of trials so that the six blocks of experimental trials took the form RA : RS : RA : RS : RA : RS. Conversely, the other group of observers, Order RS/RA, performed the first practice block under instructional set RS and the second practice block under instructional set RA, with instructions alternating in the remaining blocks. Therefore, the six experimental blocks took the form RS : RA : RS : RA : RS : RA. For both groups of observers, the first two blocks of experimental trials were treated as practice and only the last four blocks were considered in the analyses.

Practice blocks were separated by short rest periods in which observers were given the opportunity to clarify instructions. Experimental trials were performed as an unbroken sequence of 600 trials with the transitions between blocks marked only by the buzzer sounding for 1.5 sec and the concomitant change in the colour of the lamp which was illuminated.

(e) Procedure

The general procedure was identical to that used in Experiment 3 and described in Chapter 4, section A.2(e).

In addition, the observer was instructed that, on those trials for which the green lamp was illuminated she should respond as accurately as possible when the right line was longer and as quickly as possible when the left line was longer. This was instructional set RA. Alternatively, when the red lamp was illuminated she should respond as quickly as possible when the right line was longer and as accurately as possible when the left line was longer. This was instructional set RS. It was explained that one lamp would remain on over a series of trials and that there would be a pause in the stimulus sequence, marked by the sounding of a buzzer, to indicate that the alternative lamp was about to be illuminated. Due to the relative ease with which unpractised observers make accurate rather than fast responses, the speed stress of instructions was emphasised in initial instructions and during practice trials.

3. Results

Analyses of variance with Order (RA/RS or RS/RA, as defined in section C.2(d)), Period (first two experimental blocks or second two experimental blocks), Set (instructional set RA or set RS, as defined in section C.2(d)), Half (first or second half of a block), Discriminability (five levels, as defined in Chapter 4, section A.2(a)) and Response (right or left) as factors were used to analyse data from the last four experimental blocks. Analyses of variance were carried out according to a crossed factorial design with repeated measures on Period, Set, Half, Discriminability and Response. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses.

Since the Set x Response interaction, showing the effects of instructions, was significant only for response time measures, while many higher order interactions with Discriminability did reach significance and were difficult to interpret, the results of the analyses of accuracy, response time and confidence are summarised in table 23. The table shows those effects which reached significance and their probabilities. Complete analysis of variance tables for each response measure, showing the values of the F-ratios, the degrees of freedom and the error terms for these significant effects are given in Appendix 8.

SIGNIFICANT EFFECTS	ACCURACY	RESPONSE TIME	CONFIDENCE
D	< .001	< .001	< .001
P			< .05
R			< .05
S x O	< .05		
S x D		< .05	
S x R		< .01	
D x R	< .01		
P x D			< .01
S x D x O		< .01	
H x D x R	< .01		
P x H x D	< .05		
S x H x D x O		< .01	
P x H x D x R	< .01		
P x S x D x O	< .05		
P x D x R x O		< .05	
H x D x R x O	< .01	< .01	
P x S x H x D x R x O			< .05

TABLE 23. The probability values of significant main effects and interactions from the analyses of accuracy, response time and confidence. Analyses of variance were performed with Order (O), Period (P), Set (S), Half (H), Discriminability (D) and Response (R), as factors.

(a) Accuracy

The Discriminability main effect arose because the percentage of stimuli for which responses were correct increased with stimulus difference, but at a decreasing rate. The percentages of stimuli for which responses were correct at Discriminability levels 1 to 5 were 81.01%, 94.25%, 97.69%, 98.94% and 99.38%, respectively.

From table 24 it can be seen that the Set x Order interaction arose because, for Order RA/RS, accuracy was higher under Set RA than under Set RS while, for Order RS/RA, accuracy was higher under Set RS than under Set RA.

	ORDER	
SET	RA/RS	RS/RA
RA	94.70	93.40
RS	93.95	95.10

TABLE 24. Percentages of stimuli for which responses were correct in blocks of Set RA and Set RS in Order RA/RS and in Order RS/RA. The percentages were taken over the 10 observers in each order.

From table 25 it is apparent that the Discriminability x Response interaction arose because the accuracy of right and left responses differed only at Discriminability level 1, where responses to right stimuli were significantly more accurate than responses to left stimuli.

Figure 56 shows that the Half x Discriminability x Response interaction was significant because this difference in accuracy for right and left responses at Discriminability level 1 only reached significance in the first half of blocks. There was no difference in accuracy for right and left responses in the second half of blocks.

	DISCRIMINABILITY				
RESPONSE	1	2	3	4	5
Right	84.13	94.00	97.13	98.88	99.50
Left	77.88	94.50	98.25	99.00	99.25

TABLE 25. Percentages of right and left stimuli for which responses were correct at each level of discriminability. The percentages were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

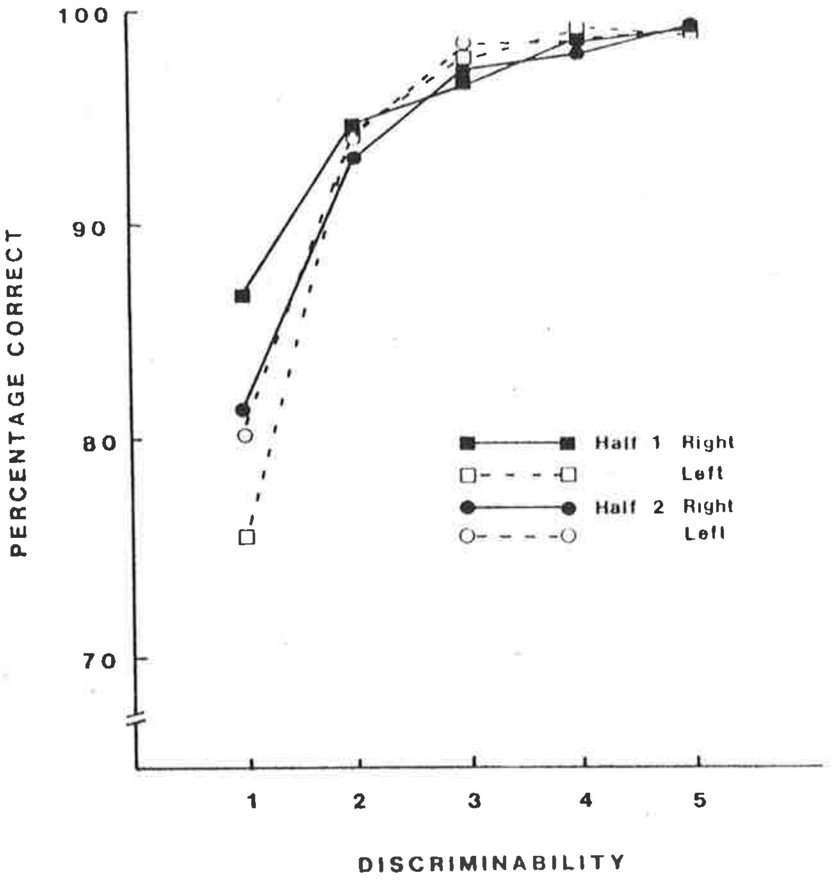


FIGURE 56. The percentages of right stimuli (solid line) and left stimuli (broken line) for which responses were correct at each level of discriminability in Half 1 (squares) and Half 2 (circles) of blocks. The percentages were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

Figure 57 shows the Period x Half x Discriminability interaction. This interaction arose because, at Discriminability level 1, there were significant differences in accuracy between the first and second halves of blocks in each period. In particular, in Period 1, accuracy increased significantly from the first to the second half of blocks while in Period 2 accuracy decreased significantly from the first to the second half of blocks. Accuracy did not vary from the first to the second half of blocks at any other time.

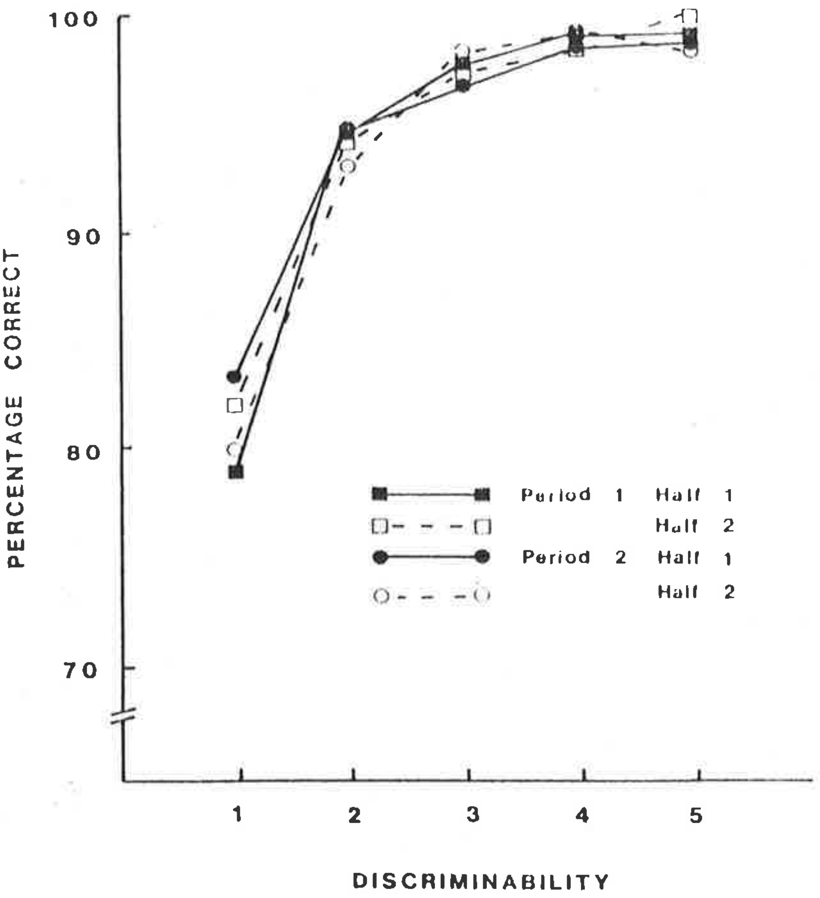


FIGURE 57. The percentages of stimuli for which responses were correct at each level of discriminability in Half 1 (solid line) and Half 2 (broken line) of blocks in Period 1 (squares) and in Period 2 (circles). The percentages were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

From table 26 it can be seen that the Period x Half x Discriminability x Response interaction arose because the difference in accuracy for right and left responses in the two halves of blocks at Discriminability level 1 also varied from Period 1 to Period 2. In particular, at Discriminability level 1 in the second half of blocks in Period 1, left responses were significantly more accurate than right responses. At all other times, at Discriminability level 1 right responses were significantly more accurate than left responses.

	PERIOD 1				PERIOD 2			
	HALF 1		HALF 2		HALF 1		HALF 2	
RESPONSE	L	R	L	R	L	R	L	R
DISCRIMINABILITY								
1	71.29	87.31	83.51	80.47	80.00	86.46	76.81	82.58
2	94.48	95.02	95.00	93.55	94.86	94.50	93.46	93.27
3	98.52	96.95	97.51	97.54	97.46	96.54	99.51	97.50
4	99.46	99.20	99.00	98.49	99.41	98.49	98.53	99.48
5	98.97	99.54	100.00	100.00	98.49	99.51	98.47	99.00

TABLE 26. Percentages of right and left stimuli for which responses were correct at each level of discriminability in Half 1 and Half 2 of blocks in Period 1 and in Period 2. The percentages were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

Similarly, from table 27 it can be seen that the difference in accuracy for right and left responses at Discriminability level 1 varied between the first and second half of blocks, depending on the order condition. In particular, there was no difference in accuracy for right and left responses at Discriminability level 1 in the second half of blocks in Order RA/RS, while in all other cases at Discriminability level 1, right responses were significantly more accurate than left responses. This pattern gave rise to the Half x Discriminability x Response x Order interaction.

	ORDER RA/RS				ORDER RS/RA			
	HALF 1		HALF 2		HALF 1		HALF 2	
RESPONSE	L	R	L	R	L	R	L	R
DISCRIMINABILITY								
1	69.13	87.00	80.01	78.30	82.06	86.49	80.50	85.21
2	96.20	94.50	94.49	94.10	93.47	95.32	94.00	92.47
3	99.13	97.49	98.37	97.52	97.25	96.41	99.12	97.53
4	99.58	99.20	99.52	99.56	99.41	98.52	98.31	98.49
5	99.50	99.46	99.50	100.00	99.03	99.48	99.00	99.30

TABLE 27. Percentages of right and left stimuli for which responses were correct at each level of discriminability in Half 1 and Half 2 of blocks in Order RA/RS and Order RS/RA. The percentages were taken over the 10 observers in each order.

From table 28 it appears that the Period x Set x Discriminability x Order interaction arose because of unsystematic changes in accuracy as stimulus difference increased. In particular, in the block of Set RS in Period 2 of Order RS/RA, accuracy was higher at Discriminability level 3 than at Discriminability levels 4 and 5.

	ORDER RA/RS				ORDER RS/RA			
	PERIOD 1		PERIOD 2		PERIOD 1		PERIOD 2	
SET	RA	RS	RA	RS	RA	RS	RA	RS
DISCRIMINABILITY								
1	78.39	78.61	82.00	74.98	83.50	81.46	82.00	87.09
2	95.98	94.54	95.46	92.85	93.26	94.00	94.00	93.68
3	98.03	97.91	98.50	97.41	96.61	97.28	95.56	100.00
4	99.50	98.60	99.36	100.00	98.48	99.53	98.00	98.51
5	99.50	100.00	100.00	98.98	99.50	99.59	98.50	98.63

TABLE 28. Percentages of stimuli for which responses were correct at each level of discriminability in blocks of Set RA and of Set RS in Periods 1 and 2. Order RA/RS and Order RS/RA are shown separately. The percentages were taken over the 10 observers in each order.

(b) <u>Response time</u>

The Discriminability main effect arose because response time decreased as stimulus difference increased, but at a decreasing rate. The mean response times at Discriminability levels 1 to 5 were 1070.55 msec, 948.04 msec, 848.03 msec, 821.31 msec and 805.87 msec, respectively.

From table 29 it is apparent that the Set x Discriminability interaction arose because, at Discriminability levels 1, 2 and 3 times for responses made under Set RA were significantly shorter than times for responses made under Set RS. In contrast, at Discriminability level 4 there was no difference in times for responses made under Sets RA and RS while at Discriminability level 5, responses made under Set RA were significantly slower than responses made under Set RS.

SET	DISCRIMINABILITY 1	2	3	4	5
RA	1050.15	944.40	840.50	822.25	813.59
RS	1090.95	951.67	855.55	820.36	798.15

TABLE 29. Mean times for correct responses at each level of discriminability in blocks of Set RA and Set RS. The means were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

The Set x Response interaction, shown in table 30, arose because, under Set RA, right responses were significantly slower than left responses, while under Set RS, left responses were significantly slower than right responses.

SET	RESPONSE LEFT	RIGHT
RA	853.40	974.61
RS	934.96	832.06

TABLE 30. Mean times for correct right and left responses in blocks of Set RA and Set RS. The means were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

From figure 58 it can be seen that the Set x Discriminability x Order interaction arose because there was no difference in mean times for responses at Discriminability levels 3, 4 and 5 in blocks performed under Set RA in Order RA/RS. In all other cases, mean response times decreased significantly from Discriminability level 3 to Discriminability level 5.

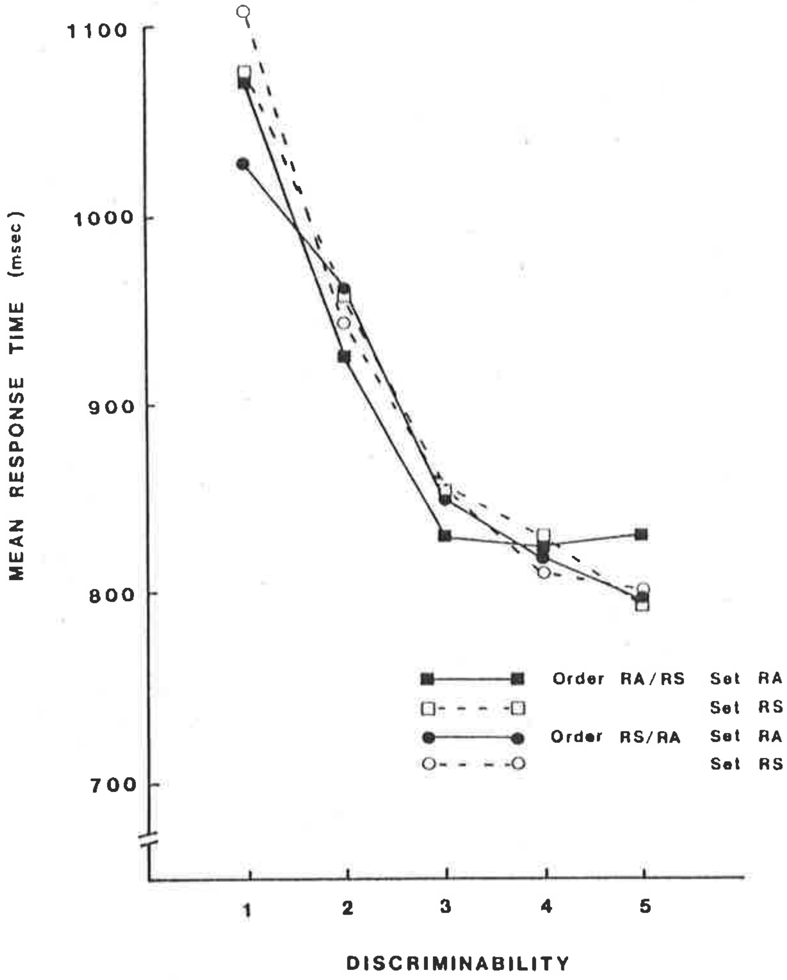


FIGURE 58. Mean times for correct responses at each level of discriminability in blocks of Set RA (solid line) and Set RS (broken line) in Order RA/RS (squares) and in Order RS/RA (circles). The means were taken over the 10 observers in each order.

Table 31 shows the Set x Half x Discriminability x Order interaction which arose because, in the first half of blocks, mean response times decreased as stimulus difference increased. In contrast, in the second half of blocks, except for blocks of Set RS in Order RS/RA, mean response times were longer at Discriminability level 5 than at Discriminability level 4.

	ORDER RA/RS				ORDER RS/RA			
	SET RA		SET RS		SET RA		SET RS	
HALF	1	2	1	2	1	2	1	2
DISCRIMINABILITY								
1	1075.55	1068.45	1047.40	1100.36	1008.08	1048.81	1169.81	1046.20
2	919.49	935.89	952.22	964.65	973.17	949.04	939.22	950.58
3	825.85	834.77	961.57	848.71	846.80	854.59	856.30	855.62
4	823.88	826.05	858.39	797.58	847.98	791.08	812.38	813.07
5	811.64	848.96	782.44	807.15	781.17	812.56	799.86	803.14

TABLE 31. Mean times for correct responses at each level of discriminability in Half 1 and Half 2 of blocks of Set RA and Set RS for Order RA/RS and for Order RS/RA. The means were taken over the 10 observers in each order.

The Period x Discriminability x Response x Order interaction is shown in table 32. It appears that this interaction arose from unsystematic variations in the mean times for right and left responses between the different levels of Discriminability in Periods 1 and 2 of each order condition. Similarly, there did not appear to be any clear pattern of variation in response time to account for the Half x Discriminability x Response x Order interaction which is shown in table 33.

	ORDER RA/RS				ORDER RS/RA			
	PERIOD 1		PERIOD 2		PERIOD 1		PERIOD 2	
RESPONSE	L	R	L	R	L	R	L	R
DISCRIMIN-ABILITY								
1	1110.38	1047.26	1068.10	1065.73	1046.87	1110.85	1087.93	1027.25
2	979.48	951.40	936.21	905.19	992.92	947.28	938.55	933.27
3	874.14	828.25	832.83	835.70	903.36	827.96	845.31	836.69
4	834.46	799.25	862.27	809.93	817.55	822.51	835.34	789.13
5	836.95	797.75	836.50	778.99	829.70	777.82	811.23	778.00

TABLE 32. Mean times for correct right and left responses at each level of discriminability in Periods 1 and 2 of Order RA/RS and of Order RS/RA. The means were taken over the 10 observers in each order.

	ORDER RA/RS				ORDER RS/RA			
	HALF 1		HALF 2		HALF 1		HALF 2	
RESPONSE	L	R	L	R	L	R	L	R
DISCRIMIN-ABILITY								
1	1112.38	1010.28	1066.11	1102.71	1080.22	1097.67	1054.58	1040.43
2	943.70	928.02	971.98	928.56	989.40	922.99	942.07	957.55
3	861.84	825.59	845.13	838.36	884.90	818.20	863.77	846.45
4	880.29	801.98	816.44	807.19	839.51	820.86	813.38	790.78
5	817.29	776.78	856.16	799.96	803.46	777.59	837.47	778.23

TABLE 33. Mean times for correct right and left responses at each level of discriminability in Half 1 and Half 2 of blocks in Order RA/RS and in Order RS/RA. The means were taken over the 10 observers in each order.

(c) Confidence

The Period main effect arose because confidence increased from Period 1 to Period 2. The mean confidence ratings for Periods 1 and 2 were 2.89 and 2.96, respectively.

The main effect of Discriminability reached significance because confidence increased with stimulus difference, but at a decreasing rate. The mean confidence at Discriminability levels 1 to 5 was 2.45, 2.75, 3.02, 3.17 and 3.24, respectively.

The Response main effect arose because confidence was higher for right responses than for left responses. The mean confidence for right and left responses was 3.04 and 2.82, respectively.

The Period x Discriminability interaction is shown in figure 59 where it can be seen that the interaction arose because, for Discriminability levels 1 and 2, confidence was significantly lower in Period 1 than in Period 2. There was no difference in confidence between the periods at Discriminability levels 3, 4 or 5.

The interaction of Period x Set x Half x Discriminability x Response x Order is shown in table 34. It appears that there was no evidence of any systematic variation in the data to give rise to this interaction.

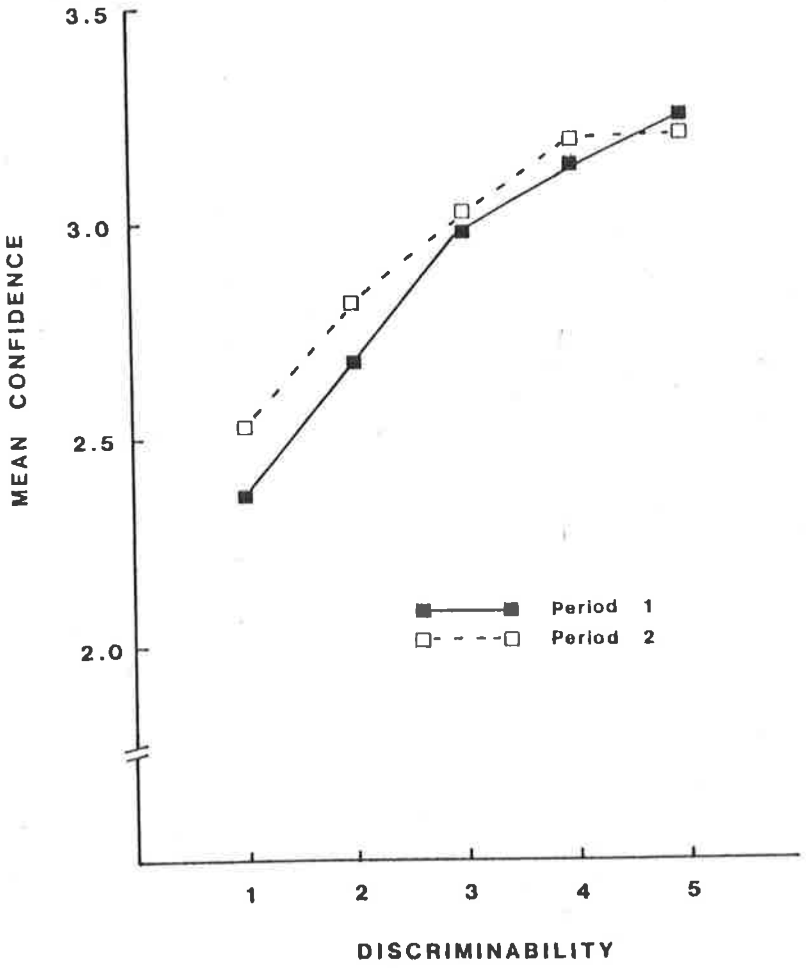


FIGURE 59. Mean confidence in correct responses at each level of discriminability in Period 1 (solid line) and Period 2 (broken line). The means were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

			PERIOD 1				PERIOD 2			
			SET RA		SET RS		SET RA		SET RS	
		HALF	1	2	1	2	1	2	1	2
ORDER RA/RS	RIGHT									
		DISCRIMINABILITY 1	2.61	2.35	2.70	2.59	2.82	2.66	2.84	2.65
		2	2.92	2.89	2.75	3.01	2.97	2.96	2.98	2.94
		3	3.02	3.14	3.11	3.24	3.21	3.37	3.12	3.16
		4	3.16	3.32	2.92	3.28	3.12	3.37	3.38	3.37
		5	3.31	3.52	3.22	3.30	3.15	3.42	3.35	3.34
	LEFT									
		1	2.30	2.59	2.60	2.54	2.52	2.64	2.49	2.34
		2	2.47	2.68	2.80	2.55	2.58	2.79	2.71	2.85
		3	2.82	3.04	2.82	2.76	2.91	2.86	2.70	2.71
		4	2.90	3.07	3.19	3.09	3.00	2.96	2.92	3.12
		5	2.98	2.74	3.18	3.12	2.76	3.27	3.21	2.95
ORDER RS/RA	RIGHT									
		DISCRIMINABILITY 1	2.29	2.38	2.25	2.27	2.45	2.63	2.41	2.56
		2	2.79	2.60	2.72	2.65	3.07	2.63	2.88	2.71
		3	3.11	3.32	3.07	3.22	3.22	3.10	3.12	3.28
		4	3.41	3.31	3.16	3.14	3.43	3.41	3.12	3.24
		5	3.40	3.50	3.44	3.58	3.41	3.51	3.34	3.37
	LEFT									
		1	2.26	2.18	1.96	2.09	2.36	2.15	2.45	2.44
		2	2.48	2.50	2.51	2.55	2.51	2.89	2.78	2.81
		3	2.70	2.80	2.85	2.84	2.78	3.16	3.04	3.10
		4	3.06	3.12	3.08	3.02	2.80	3.50	3.23	3.23
		5	3.23	3.16	3.20	3.08	3.00	3.02	3.05	3.31

TABLE 34. Mean confidence in correct right and left responses at each level of discriminability in each half of blocks of Set RA and of Set RS. The means are shown for Periods 1 and 2 of Order RA/RS and of Order RS/RA. The means were taken over the 10 observers in each order.

(d) Confidence and response time

For each block of trials, confidence ratings for both correct and incorrect responses were available for each of the 20 observers for each level of discriminability and for right and left responses. These ratings were plotted against the corresponding response time values and straight lines were fitted to each of these 800 plots. It was found that 602 of the slopes of the best fitting straight lines were negative, showing that there was a predominantly inverse relationship between confidence and response time within blocks ($p < .001$, binomial test). This inverse relationship was still evident when the 400 plots for Set RA and for Set RS were considered separately. In this case,

DISCRIMINABILITY	SET RA No. of negative slopes	p	SET RS No. of negative slopes	p	OVERALL No. of negative slopes	p
1	70	< .001	76	< .001	146	< .001
2	72	< .001	76	< .001	148	< .001
3	56	< .001	60	< .001	116	< .001
4	60	< .001	70	< .001	130	< .001
5	62	< .001	52	< .001	114	< .001

TABLE 35. The number of negative slopes and the associated binomial probabilities obtained when, for each of the 20 observers, at each level of discriminability, confidence ratings were plotted against the corresponding response time values for right and left responses separately, and straight lines were fitted. Data from each block performed under each instructional set were considered separately, yielding a total of 80 plots for each set, or 160 plots overall.

the numbers of negative slopes were 320 and 334, respectively ($p < .001$, $p < .001$, respectively, binomial test). Therefore, there was a highly significant inverse relationship between confidence and response time within each block, irrespective of the instructional set. When individual levels of discriminability were considered separately, combining data from both instructional sets at each level, a significantly higher proportion of the 160 slopes for each level was negative at all levels of discriminability. This was also the case when the 80 slopes for Set RA and for Set RS were considered separately for each level of discriminability. In this case, the actual proportions with the associated probabilities are given in table 35.

4. Discussion

The majority of the effects which reached significance did not appear to reflect the manipulation of relative caution by variations in instructions. Instead, these findings appeared to arise from unsystematic changes in the response measures between the levels of discriminability in the different order conditions. As a consequence, although these findings may not be inconsistent with any of the decision processes, they do not correspond to simple predictions of the processes. Therefore, they will not be discussed in detail. In particular, due to the difficulty of interpretation of higher order interactions, and since these findings are not a simple result of the experimental manipulation, only main effects and 2-way and 3-way interactions will be discussed.

The only finding which indicated that instructions produced changes in relative caution was the Set x Response interaction in response time data. This effect showed that observers complied with instructions to make one response more quickly than the other and therefore it is consistent with the predictions of both the ideal observer hypothesis and the target confidence mechanism. However, the lack of any equivalent variation in accuracy or confidence contrasts with the predictions of all of the processes.

As in earlier experiments in this thesis, accuracy and confidence increased and response time decreased as stimulus difference increased. The rate of change in all three response measures decreased at large values of stimulus difference. This Discriminability effect is consistent with previous research and it is predicted by all of the decision processes. At the same time, as in previous studies, there was a clear inverse relationship between response time and confidence within blocks of trials, both when all levels of discriminability were considered together and when each was considered separately. This finding is consistent with Audley's (1960) hypothesis and the balance of evidence hypothesis.

(a) Accuracy

The Set x Order interaction in accuracy data would appear to indicate that observers were more accurate when performing under the instructional set which they experienced first. This could indicate that accuracy was decreasing from the beginning to the end of the experimental session as a result of fatigue. However, this seems to be a less plausible explanation when it is noted that the Period effect was not significant. Alternatively, it is possible that observers found the task less difficult when performing under the set that they learnt first, resulting in higher accuracy. It is possible that the adoption of one instructional set is not very difficult but that the subsequent adoption of an alternative, very similar set leads to some confusion in responding, with a consequent reduction in accuracy.

The Discriminability x Response interaction in accuracy data may be explained in terms of a tendency to favour right responses over and above any variation in relative caution induced by instructions. This preference for right responses overrode instructions only at the smallest stimulus difference. A preference for right responses does not seem unlikely in right-handed observers and it has been observed in previous experiments in the thesis where it occurred at all levels of Discriminability. The finding that, in this case, the effect was restricted to Discriminability level 1 appears to indicate the influence of averaging the two sets, one of which, instructional Set RA, operates against a preference for right responses by stressing left responses. As a result, the preference for right responses would be apparent only when it was strongest, at Discriminability level 1.

Variations in the tendency to favour right responses as the experimental session progressed and between order conditions would appear to account for the interactions of Period x Half x Discriminability x Response and Half x Discriminability x Response x Order, while the finding that the pre-existing preference for right responses is only significant at Discriminability level 1 in the first half of blocks gives rise to the Half x Discriminability x Response interaction.

The Period x Half x Discriminability interaction does not appear to be directly related to the manipulation of instructions. The increase in accuracy from Half 1 to Half 2 of blocks in Period 1 may indicate the influence of increasing familiarity with the requirements of the instructional set as the block progressed. In contrast, the decrease in accuracy from Half 1 to Half 2 of blocks in Period 2 appears to indicate a decline in performance due to fatigue at the end of blocks of trials which are performed at the conclusion of a lengthy experimental session.

(b) Response time

Many of the significant effects in response time data appear to arise from an increase in the variability in response times at the larger values of stimulus difference. For example, the Set x Discriminability interaction arose because, at Discriminability levels 1, 2 and 3, response times under Set RA were faster than response times under Set RS. This appears to result from the combination of a natural tendency to favour right responses with the different instructional sets. Since right responses tended to be faster than left responses, irrespective of instructions, it would be expected that when instructions stressed the speed of left responses, as in Set RA, overall response times would be faster. However, this trend was not found at Discriminability levels 4 and 5. There was no difference in response times under the two sets at Discriminability level 4 and responses under Set RA were slower than responses under Set RS at Discriminability level 5. Although there is no apparent reason for this variation, it does correspond to the increase in the variability of response times at the largest stimulus difference which has been found in previous studies in the thesis.

Similarly, the Set x Discriminability x Order interaction occurs because the inverse relationship between response time and discriminability did not always hold at large values of stimulus difference. In particular, there was no difference in response times at Discriminability levels 3, 4 and 5 in blocks of Set RA in Order RA/RS. In general, it appears that the unsystematic changes in response time occurred because, at the larger values of stimulus difference, times were tending to a minimum. Therefore, at this point, relatively small variations would be noticeable and would be likely to give rise to significant effects.

(c) Confidence

The Period main effect in confidence data appears to indicate the effects of fatigue since confidence decreased from Period 1 to Period 2. In addition, the finding that this effect was more marked at Discriminability levels 1 and 2 produced the Period x Discriminability interaction. Fatigue effects are not unlikely in a lengthy experimental session and are not inconsistent with any of the decision processes. But, if this effect is due to fatigue, it is surprising that it was evident only in confidence data.

As suggested by accuracy data, the Response effect in confidence appears to indicate a tendency to favour right responses irrespective of instructions.

5. Summary and Conclusions

In this experiment, only response time measures showed any systematic variation with changes in instructions. The observed changes in response time were consistent with the predictions of both the ideal observer hypothesis and the target confidence mechanism. Since there were no effects on confidence data no distinction can be drawn between the processes. The majority of the significant effects took the form of higher order interactions with Discriminability and were best explained in terms of unsystematic changes in performance between levels of Discriminability which were not directly related to the manipulation of instructions.

D. EXPERIMENT 9

1. Introduction

The experiment was repeated with discriminability held constant as the relative caution of instructions was varied according to a within-subjects design. In order to avoid ceiling effects which were evident in accuracy data at Discriminability levels 3, 4 and 5, level 2 was used for all stimuli. The use of level 2 rather than level 1 meant that completely error-free performance, in accordance with instructions, would be possible.

2. Method

(a) Stimuli

The stimuli were white pi-shaped figures identical to those used in Experiment 3 and described in Chapter 4, section A.2(a) except that, in this case, discriminability did not vary. The longer line was always 187 mm in length while the shorter line was always 186 mm in length.

(b) Apparatus

The apparatus was identical to that used in Experiment 8 as described in section C.2(b).

(c) Observers

The 20 observers were men enrolled in the first year psychology course at the University of Adelaide, whose participation was credited towards a course requirement. They ranged in age from 17 to 20 years and all declared themselves to be right-handed. Observers had accurate red-green colour discrimination, and all were naive with respect to the aims of the experiment.

(d) Design

The design was identical to that of Experiment 8 as described in section C.2(d) except that, in this case, discriminability did not vary.

(e) Procedure

The procedure was identical to that used in Experiment 8 as described in section C.2(e).

3. Results

Analyses of variance with Order (RA/RS or RS/RA, as defined in section C.2(d)), Period (first two experimental blocks or second two experimental blocks), Set (instructional Set RA or Set RS, as defined in section C.2(d)), Half (first or second half of a block) and Response (right or left) as factors were used to analyse data from the last four experimental blocks. Analyses of variance were carried out according to a crossed factorial design with repeated measures on Period, Set, Half and Response. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses.

(a) Accuracy

The analysis of accuracy data showed a significant main effect of Half ($F(1,18) = 9.26$, $MSerror = 0.02$, $p < .01$) and a significant Set x Response interaction ($F(1,18) = 17.11$, $MSerror = 0.20$, $p < .001$).

The Half main effect arose because accuracy was lower in the first than in the second half of blocks. The percentages of stimuli for which responses were correct in Half 1 and Half 2 of blocks were 94.55% and 95.05%, respectively.

The Set x Response interaction is summarised in table 36. Under Set RA, right responses were significantly less accurate than left responses, while under Set RS, right responses were significantly more accurate than left responses.

		SET	
		RA	RS
	RESPONSE		
ACCURACY	RIGHT	92.85	96.75
	LEFT	97.05	92.55
RESPONSE TIME	RIGHT	918.28	790.67
	LEFT	820.66	983.13

TABLE 36. Percentages of right and left stimuli for which responses were correct and mean times for correct right and left responses in blocks of Set RA and of Set RS. Measures were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

(b) Response time

The analysis of response time data showed significant interactions of Period x Response (F(1,18) = 6.83, MSerror = 6146, p < .05), Set x Response (F(1,18) = 20.42, MSerror = 82410, p < .001), Period x Set x Half x Order (F(1,18) = 5.18, MSerror = 8464, p < .05) and Period x Set x Response x Order (F(1,18) = 8.07, MSerror = 4735, p < .05).

The mean times for right and left responses in Periods 1 and 2 are shown in table 37. The Period x Response interaction arose because right responses were significantly slower in Period 1 than in Period 2, while there was no difference in the times for left responses in Periods 1 and 2. Right responses were significantly faster than left responses in both periods.

	PERIOD	
RESPONSE	1	2
RIGHT	874.67	834.28
LEFT	899.19	904.61

TABLE 37. Mean times for correct right and left responses in Period 1 and in Period 2. The means were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

The Set x Response interaction, shown in table 36 arose because, under Set RA, right responses were significantly slower than left responses, while, under Set RS, right responses were significantly faster than left responses.

The Period x Set x Half x Order interaction is shown in figure 60 where it can be seen that the change in response time from the first to the second half of blocks varied between sets and periods in the two order

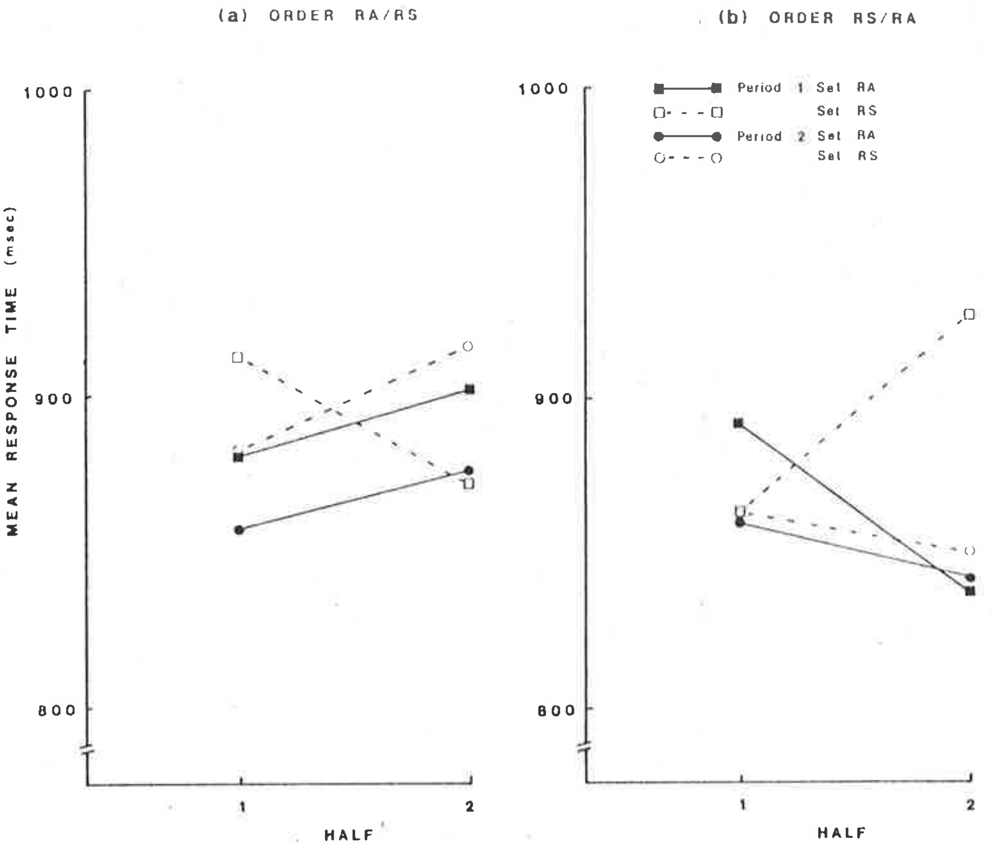


FIGURE 60. Mean times for correct responses in Half 1 and Half 2 of blocks of Set RA (solid line) and Set RS (broken line) in Period 1 (squares) and Period 2 (circles) in (a) Order RA/RS and (b) Order RS/RA. The means were taken over the 10 observers in each order.

conditions. In general, times increased from the first to the second half of blocks in Order RA/RS and decreased from the first to the second half of blocks in Order RS/RA. However, this was not always the case. In both order conditions, times for responses made under Set RS in Period 1 changed in the opposite way. In particular, in Order RA/RS, times decreased from the first to the second half of the block of Set RS in Period 1, while in Order RS/RA, times increased from the first to the second half of the block of Set RS in Period 1.

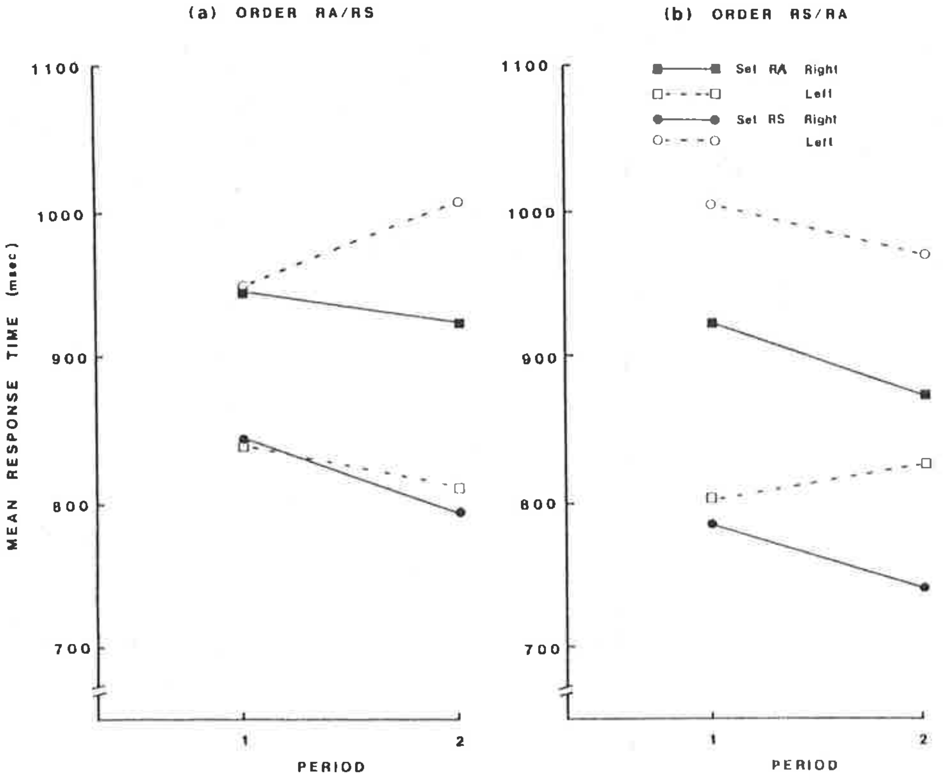


FIGURE 61. Mean times for correct right responses (solid line) and left responses (broken line) in blocks of Set RA (squares) and Set RS (circles) in each period of (a) Order RA/RS and (b) Order RS/RA. The means were taken over the 10 observers in each order.

From figure 61 it is apparent that the Period x Set x Response x Order interaction was significant because right responses showed a regular pattern of variation while the pattern for left responses varied between sets within each order condition. For both orders and under both sets, right responses were slower in Period 1 than in Period 2. In contrast, in Set RA of Order RA/RS and Set RS of Order RS/RA, left responses were slower in Period 1 than in Period 2, while in Set RA of Order RA/RS and Set RS of Order RS/RA, left responses were faster in Period 1 than in Period 2.

(c) Confidence

The analysis of confidence data showed significant interactions of Set x Order (F(1,18) = 4.97, MSerror = 0.15, $p < .05$), Period x Response (F(1,18) = 6.46, MSerror = 0.10, $p < .05$), Period x Half x Response x Order (F(1,18) = 10.48, MSerror = 0.07, $p < .01$) and Period x Set x Half x Response (F(1,18) = 5.72, MSerror = 0.08, $p < .05$).

Table 38 shows that the Set x Order interaction arose because, for both sets, confidence was lower for Order RA/RS than for Order RS/RA. At the same time, for Order RA/RS, confidence was lower in Set RA than in Set RS, while for Order RS/RA, confidence was higher in Set RA than in Set RS.

	SET	
ORDER	RA	RS
RA/RS	2.99	3.10
RS/RA	3.50	3.42

TABLE 38. Mean confidence in correct responses in blocks of Set RA and Set RS in Order RA/RS and in Order RS/RA. The means were taken over the 10 observers in each order.

The Period x Response interaction, shown in table 39, was significant because there was a significant increase in confidence in right responses from Period 1 to Period 2 while confidence in left responses did not change.

Figure 62 shows the Period x Half x Response x Order interaction. In Order RA/RS, there was a significant decrease in confidence in left responses from Half 1 to Half 2 of blocks in Period 2. Confidence did not

RESPONSE	PERIOD 1	PERIOD 2
RIGHT	3.20	3.43
LEFT	3.18	3.22

TABLE 39. Mean confidence in correct right and left responses in Period 1 and Period 2. The means were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

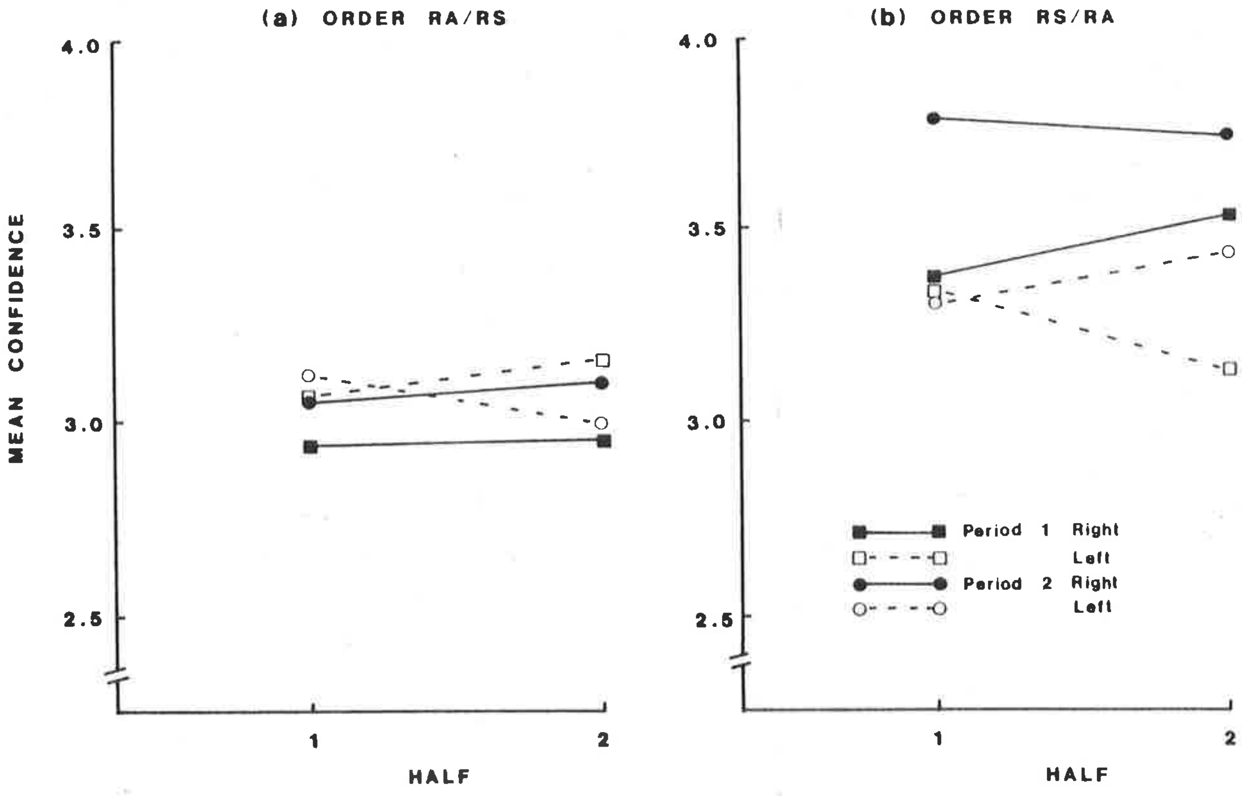


FIGURE 62. Mean confidence in correct right responses (solid line) and left responses (broken line) in Half 1 and Half 2 of blocks in Period 1 (squares) and Period 2 (circles) of (a) Order RA/RS and (b) Order RS/RA. The means were taken over the 10 observers in each order.

change between the halves of blocks at any other time in Order RA/RS. In Period 1 of Order RS/RA, confidence in right responses increased significantly and confidence in left responses decreased significantly from Half 1 to Half 2. In Period 2 of Order RS/RA, confidence in right responses did not change between halves, but confidence in left responses increased significantly from Half 1 to Half 2.

The Period x Set x Half x Response interaction is shown in figure 63 where it can be seen that, under both sets, confidence in right responses in

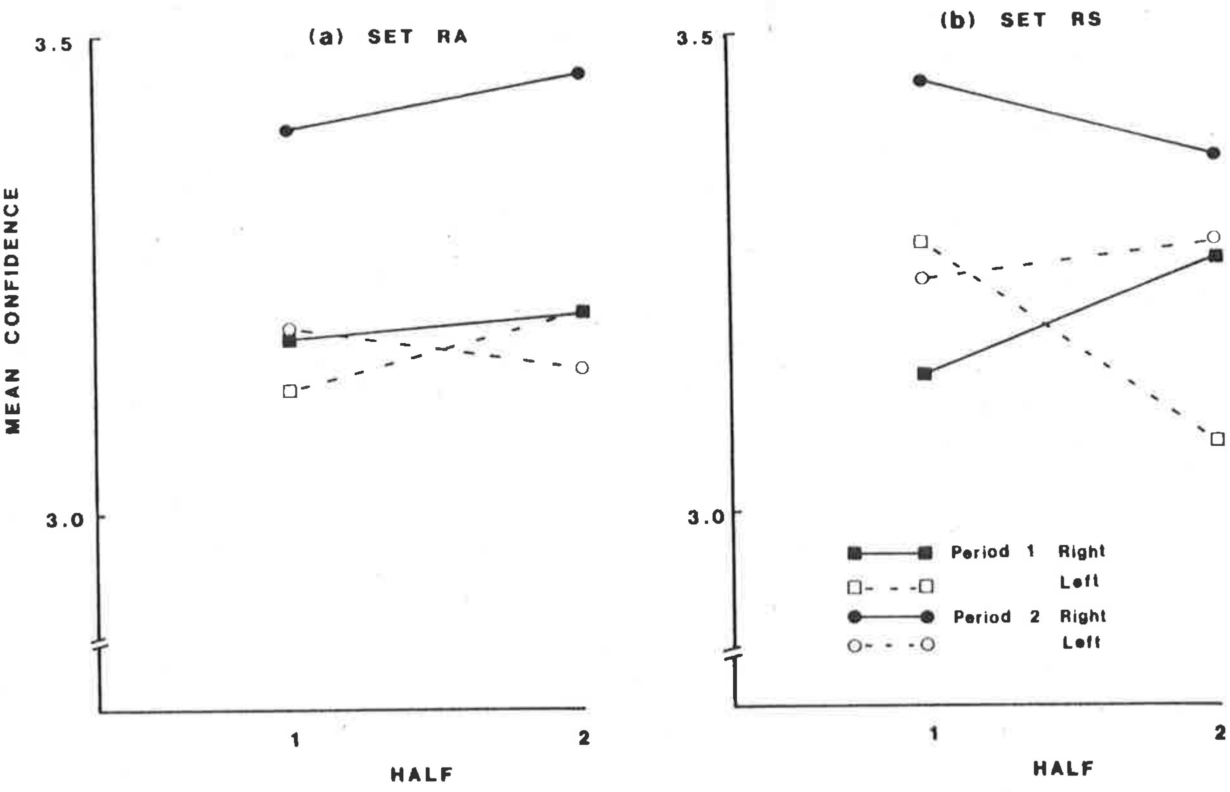


FIGURE 63. Mean confidence in correct right responses (solid line) and left responses (broken line) in Half 1 and Half 2 of blocks in Period 1 (squares) and Period 2 (circles) of (a) Set RA and (b) Set RS. The means were taken over 20 observers, the 10 observers in Order RA/RS and the 10 observers in Order RS/RA.

Period 1 was higher than both confidence in left responses in Period 1 and confidence in right and left responses in Period 2. There were no significant differences in confidence between Half 1 and Half 2 except for left responses in Period 1 of Set RS where there was a significant decrease from Half 1 to Half 2.

(d) Confidence and response time

For each block of trials, confidence ratings were available for each of the 20 observers for both right and left responses. These ratings were plotted against the corresponding response time values and straight lines were fitted to each of these 160 plots. It was found that 141 of the slopes of the best fitting straight lines were negative, showing that there was a predominantly inverse relationship between confidence and response time within blocks of trials ($p < .001$, binomial test). This inverse relationship was still evident when the 80 plots for each instructional set were considered

	RESPONSE	NO. OF NEGATIVE SLOPES	p (binomial test)
SET RA	RIGHT	33	<.001
	LEFT	38	<.001
SET RS	RIGHT	35	<.001
	LEFT	35	<.001

TABLE 40. The number of negative slopes and the associated binomial probabilities obtained when, for each of the 20 observers, confidence ratings were plotted against the corresponding response time values for right and left responses separately, and straight lines were fitted. Data from each block performed under each instructional set were considered separately, yielding a total of 40 plots for each response in each set.

separately. In this case, the numbers of negative slopes for Set RA and for Set RS were 71 and 70, respectively ($p < .001$ in both cases, binomial test). Thus, there was a significant inverse relationship between confidence and response time within each block of trials and irrespective of the instructional set. When right and left responses were considered separately for Set RA and Set RS, the relationship was still significant for each response under each instructional set. The number of negative slopes from the 40 plots available for each of right and left responses under the two sets of instructions is shown in table 40 where the associated probabilities are given.

4. Discussion

The changes in relative caution due to instructions were evident in the Set x Response interaction which accounted for a significant proportion of the variance in both accuracy and response time data. The observed changes in accuracy and response time are consistent with the predictions of both the ideal observer hypothesis and the target confidence mechanism. However, since there was no significant change in confidence in response to variations in instructions it is not possible to make any distinction between the processes.

The lack of variation in confidence with changes in instructional set may be due to the emphasis on speed rather than accuracy of responding in the original instructions and during practice trials. This emphasis may have given rise to the relatively small differences in mean response times between Set RA and Set RS of about 100-150 msec. Therefore, the longer times under Set RS are likely to overlap with the shorter times under Set RA. Since it would be expected that confidence ratings would be related to response times, independent of the instructional set, this would result in a large amount of overlap between the confidence ratings from the different sets.

Although there was no systematic change in confidence between instructional sets there was a clear inverse relationship between confidence and response time within each set. This relationship, which is consistent with all of the formulations for confidence, has been observed in previous experiments and perseveres across variations in instructions or stimulus properties.

The other effects which reached significance did not appear to arise from the manipulation of relative caution by varying instructions. For example, in accuracy data, the Half main effect does not obviously reflect an instruction induced change in relative caution. Higher accuracy in the first half of blocks might indicate that observers are more attentive immediately after a change in instructions.

Similarly, response time and confidence data showed several interactions which cannot be explained in terms of the manipulation of instructions. For example, the Period x Response interaction found in both response time and confidence measures did not arise from the variation in instructions. Indeed, this interaction does not appear to be consistent with any explanation in terms of experimental factors. It is possible that this effect arose from variations in response time as the experimental session progressed which were unrelated to the experimental manipulation and which, in turn, gave rise to variations in confidence.

The Period x Set x Half x Order interaction in response time data arose because, from Half 1 to Half 2 in the first block of Set RS in each order condition, times varied in a direction opposite to the variation in any other block. There appears to be no possible explanation of this effect in terms of the changes in instructions. Once more it appears that this change in response times is not related to the experimental manipulation. Similarly, the Period x Set x Response x Order interaction in response time data arose because of unsystematic changes in times for left responses between periods.

In the second type of block performed in each order condition, times for left responses increased from Period 1 to Period 2 while, in all other cases, times for right and left responses decreased from Period 1 to Period 2. The reason for this variation is not apparent and does not appear to involve the manipulation of instructional set.

The Set x Order interaction in confidence data can only be explained in terms of random variations in confidence between the order conditions. In particular, it does not appear to result from any intended change in relative caution due to instructions. Similarly, the Period x Half x Response x Order interaction in confidence measures appears to be the result of unsystematic changes in confidence in each response from the first to the second half of blocks as the experimental session progressed. At the same time, the Period x Set x Half x Response interaction is evidence of random variations in confidence ratings for each response from the two instructional sets during the experimental session.

5. Summary and Conclusions

It appears that, as with changes in stimulus probability and discriminability, the use of instructions to change relative caution does not have a very strong influence on performance. Indeed, when discriminability was varied within each instructional set in Experiment 8, only response time showed any consistent change due to instructions. When only instructions were varied in Experiment 9, both accuracy and response time showed consistent changes due to instructions but confidence still remained unchanged. In both experiments, the changes in both accuracy and response time were consistent with the predictions of both the ideal observer hypothesis and the target confidence mechanism. However, since in this situation these mechanisms could be differentiated only on the basis of their predictions about confidence, it was not possible to distinguish between them on the basis of these data.

CHAPTER 8

A. INTRODUCTION

In Experiment 9, the manipulation of relative caution by instructions produced systematic changes in accuracy and response time but not in confidence. This may have arisen because instructions emphasised the response which was to be made quickly at the expense of stressing the accuracy of the alternative response. It appears that this resulted in very similar times for both responses under both sets of instructions. Since it seems likely that confidence is related to response time, the difference between confidence for each response under the two sets of instructions would be expected to be correspondingly small. The differentiation between the levels of caution may be enhanced if instructions manipulate overall rather than relative caution, so that speed is not emphasised at the expense of accuracy.

As seen in Chapter 1, variations in caution have been studied using choice reaction time tasks (e.g. Howell & Kreidler, 1963; Fitts, 1966; Pachella & Pew, 1968; Hale, 1969) and same-different judgments (e.g. Link, 1971; Link & Tindall, 1971; Ascher, 1974). However, only four studies are known in which there has been an explicit manipulation of caution in a two-category discrimination task. These studies by Garrett (1922), Johnson (1939), Festinger (1943a) and Wilding (1974) were described in detail in Chapter 1, so only a brief summary of the findings will be presented here.

In these studies, caution was varied between separate experimental sessions either by changing the stimulus display time or by the use of instructions. Garrett (1922), Johnson (1939) and Festinger (1943a) found that accuracy and response time increased with caution, while confidence did not change. On the other hand, Wilding (1974) found that response time increased while accuracy remained the same as caution increased. Confidence measures were not taken in Wilding's (1974) study, but discriminability was varied within each level of caution. He noted that, within a given level of caution, accuracy decreased and response time increased as discriminability decreased and that the rate of change of the response measures was fastest when instructions stressed accuracy so that caution was high.

The lack of any variation in confidence in the studies of Garrett (1922), Johnson (1939) and Festinger (1943a) is surprising in the light of the general finding that, within observers, confidence increases with accuracy as discriminability increases (Garrett, 1922; Johnson, 1939; Festinger, 1943a; Pierrel & Murray, 1963). Given the evidence reviewed in Chapter 1 which indicates that observers can and do trade speed of responding for accuracy, it is surprising that, when observers increase the overall accuracy of their performance, for example, in response to instructions, this increase in discriminative capacity is not reflected in confidence ratings. At the same time, the finding that confidence did not vary in a situation in which it would be assumed that criterion values are varying is not consistent with any of the decision processes.

As seen in Chapter 1, Ascher (1974) and Vickers (1979) suggested that the lack of any variation in confidence may have arisen as the result of an unwanted range effect (Poulton, 1973, 1975, 1979) due to the nature of the experimental design. Ascher (1974) and Vickers (1979) noted that, since caution was varied between separate experimental sessions, observers may have rated their confidence in judgments in relation to other judgments in that session alone, rather than in comparison with all judgments made in all experimental sessions. For example, in Chapter 1 it was noted that, in accordance with instructions to use an entire percentage scale, observers may have decided to rate their most confident response in any one session as 100, and their least confident response as 0. As a consequence, within each level of caution, confidence ratings may be spread over the entire scale.

B. PREDICTIONS OF THE DECISION PROCESSES

As noted in Chapter 7, neither the response stabilisation hypothesis nor the adaptation-level hypothesis is designed to account for changes in performance due to variations in instructions. Although the effects of instructions do not have any direct representation in the ideal observer hypothesis, in Chapter 7 it was suggested that changes in instructions would be likely to produce variations in the subjective costs and payoffs associated with each decision. Therefore, if the observer is trying to maximise the expected value of his decisions, it would be expected that the cutoff will vary as instructions change the subjective costs and payoffs. However, in the case in which instructions are used to manipulate overall caution for both responses, so that the subjective costs or payoffs associated with both responses will vary in the same way, it would be expected that the overall sensitivity of the observer, d', would vary rather than the criterion, β. Therefore, only the target confidence mechanism provides predictions about changes in performance when overall caution is manipulated by instructions.

The target confidence mechanism applied to the accumulator model predicts that criterion values will change in response to changes in instructions. For example, if instructions stress the accuracy of both responses, it would be assumed that the target confidence for both responses will be high. As a consequence, the criteria for both responses in the primary decision process will increase, giving rise to changes in all three response measures.

For example, when instructions stress accuracy and the criteria for both responses are high, responses will be made only after considerable evidence has been collected. Therefore, it is unlikely that the criterion will be reached in error due to the random accumulation of evidence in favour of the inappropriate response. As a result, accuracy will tend to be high. Since many observations will be required before either high criterion is satisfied, times for both responses will tend to be long. At the same time, the maximum possible confidence in either response will be equal to the high criterion value, so mean confidence for both responses will tend to be high. When instructions stress the overall speed of responding, the predictions will be the opposite of those just described.

When the target confidence mechanism is applied to the random walk model in the situation in which instructions stress the accuracy of both responses, the increase in the target confidence for both responses may be assumed to produce a movement of both boundaries away from the starting point. Therefore, the qualitative predictions of this process are identical to those for the accumulator model described above.

C. EXPERIMENT 10

1. Introduction

In this experiment, instructions were used to change the level of caution between successive blocks within a single experimental session. In an attempt to increase the range of confidence ratings, discriminability was varied within each level of caution.

2. Method

(a) Stimuli

The stimuli were white, pi-shaped figures, identical to those used in Experiment 3 and described in Chapter 4, section A.2(a) with the five levels of discriminability as defined therein.

(b) Apparatus

The apparatus was identical to that used in Experiment 3 and described in Chapter 4, section A.2(b), with one addition. On top of the video monitor, directly above the stimulus display, was a black board on which were positioned a red and a green neon lamp and a buzzer. The lamps were 1 cm in diameter and 15 cm apart, centre to centre. The green lamp, on the right side of the board, was labelled 'fast' and the red lamp, on the left side of the board, was labelled 'accurate'. The buzzer was placed in the centre of the board. The PDP 8/E computer controlled the illumination of either lamp and the sounding of the buzzer.

(c) Observers

The 20 observers, 10 men and 10 women, were students enrolled in the first year psychology course at the University of Adelaide, whose participation was credited towards a course requirement. They ranged in age from 18 to 25 years and all declared themselves to be right-handed. Observers had accurate red-green colour discrimination and all were naive with respect to the aims of the experiment.

(d) Design

Two groups of 10 observers were formed, each comprising 5 men and 5 women. The groups performed under different order conditions with each observer being tested individually for approximately one hour.

In both order conditions, stimuli were presented in 10 blocks: 4 practice blocks each of 50 trials preceded 6 experimental blocks each of 100 trials. For both groups of observers, and within each block of trials, both practice and experimental, the occurrence of the longer line on the right or the left was equiprobable (p(R)=0.5) and the sequence was random. In addition, each of the five levels of discriminability occurred equally often for right stimuli and for left stimuli within each block and the sequence was random. Each block of trials was performed under one of two types of instructional set:

A. Respond as accurately as possible
S. Respond as quickly as possible

For all observers, the first and third practice blocks were performed under Set S, and the second and fourth practice blocks were performed under Set A. One group of observers, Order AS, performed the first experimental block under Set A while the other group of observers, Order SA, performed the first experimental block under Set S. The instructional sets alternated in the remaining five experimental blocks. The first two experimental blocks were treated as practice, so only the last four experimental blocks were considered in the analyses.

Practice blocks were separated by short rest periods in which observers were given the opportunity to clarify instructions. Experimental blocks were performed as an unbroken sequence with the transitions between blocks marked only by the buzzer sounding for 1.5 sec and the concomitant change in the colour of the lamp which was illuminated.

(e) Procedure

The general procedure was identical to that used in Experiment 3 and described in Chapter 4, section A.2(e).

In addition, the observer was instructed that, on those trials for which the red lamp was illuminated he should respond as accurately as possible, attempting to achieve error-free performance. This was instructional Set A. Alternatively, on those trials for which the green lamp was illuminated he should respond as quickly as possible. This was instructional Set S. It was explained that one lamp would remain on over a series of trials and that there would be a pause in the stimulus sequence, marked by the sounding of the buzzer, to indicate that the alternative lamp was about to be illuminated.

3. Results

Analyses of variance with Order (AS or SA, as defined in section C.2(d)), Period (first two experimental blocks or second two experimental blocks), Set (A or S, as defined in section C.2(d)), Half (first or second half of a block), Discriminability (five levels, as defined in Chapter 4, section A.2(a)) and Response (right or left) as factors were used to analyse data from the last four experimental blocks. Analyses of variance were carried out according to a crossed factorial design with repeated measures on Period, Set, Half, Discriminability and Response. Separate analyses were performed on (a) the percentage of stimuli for which responses were correct, (b) the mean times for correct responses and (c) the mean confidence in correct responses.

The results of the analyses of accuracy, response time and confidence are summarised in table 41 which shows only those effects which reached significance and their probabilities. From the table it is apparent that only

SIGNIFICANT EFFECT	ACCURACY	RESPONSE TIME	CONFIDENCE
S	< .001	< .001	< .01
D	< .001	< .001	< .001
S x D	< .05	< .001	< .001
S x D x R	< .05		
S x H x O	< .05		
P x H x R		< .05	
P x D x R		< .01	
P x S x D			< .01
S x H x R x O		< .05	
H x D x R x O		< .05	
P x H x D x R			< .05
P x S x D x O	< .05		
P x S x H x D x O		< .05	
P x S x D x R x O		< .05	
P x S x H x R x O			< .05
S x H x D x R x O		< .05	
P x S x H x D x R x O	< .05		< .01

TABLE 41. The probability value of significant main effects and interactions from the analyses of accuracy, response time and confidence. Analyses of variance were performed with Order (O), Period (P), Set (S), Half (H), Discriminability (D) and Response (R) as factors.

the main effects of Set and Discriminability and the Set x Discriminability interaction were significant for all three response measures. At the same time, many higher order interactions with Discriminability and Order reached significance for individual response measures and were difficult to interpret. Complete analysis of variance tables for each response measure showing the values of F-ratios, the degrees of freedom and the error terms for these significant effects are given in Appendix 10.

(a) Accuracy

The Set main effect arose because accuracy was higher under Set A than under Set S. The percentages of stimuli for which responses were correct under Sets A and S were 94.28% and 85.86%, respectively.

The Discriminability main effect reached significance because accuracy increased with stimulus difference, but at a decreasing rate. The percentages of stimuli for which responses were correct at Discriminability levels 1 to 5 were 76.76%, 88.08%, 93.18%, 95.74% and 96.59%, respectively.

The interaction between Set and Discriminability is shown in figure 64 where it is apparent that accuracy rates tended to an asymptote more rapidly under Set A than under Set S. When normal ogives were fitted to these data, the standard deviations of the ogives for Set A and Set S were 2.30 and 1.19, respectively.

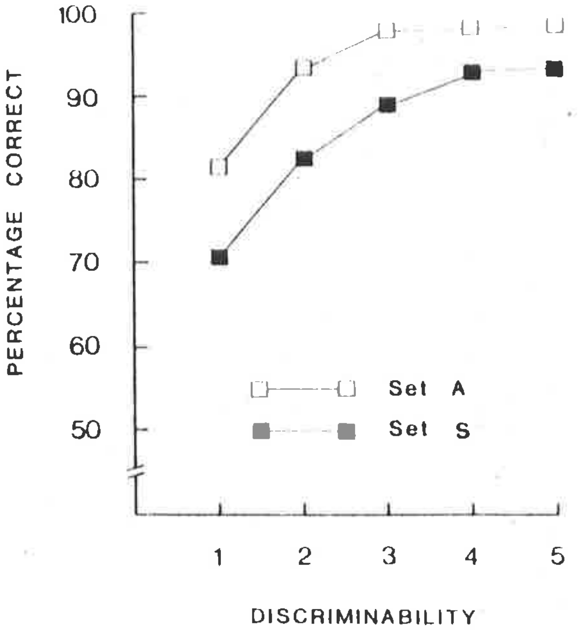


FIGURE 64. Percentage of stimuli for which responses were correct under Set A (open squares) and Set S (filled squares). The percentages were taken over 20 observers, the 10 observers in Order AS and the 10 observers in Order SA.

The Set x Discriminability x Response interaction is shown in table 42. This interaction arose because, under Set A, responses to right stimuli were significantly more accurate than responses to left stimuli at Discriminability level 1, while responses to right stimuli were significantly less accurate than responses to left stimuli at Discriminability level 2. Under Set S, responses to right stimuli were significantly less accurate than responses to left stimuli at both Discriminability levels 1 and 2. There was no significant difference in accuracy for right and left responses at any other level of Discriminability.

		SET A		SET S	
RESPONSE		LEFT	RIGHT	LEFT	RIGHT
DISCRIMINABILITY	1	80.22	83.33	75.76	67.69
	2	95.23	92.55	85.07	79.46
	3	98.75	97.73	88.36	87.88
	4	98.16	99.15	92.96	92.67
	5	98.94	98.71	94.07	94.63

TABLE 42. Percentage of right and left stimuli for which responses were correct at each level of discriminability in blocks of Set A and of Set S. The percentages were taken over 20 observers, the 10 observers in Order AS and the 10 observers in Order SA.

Table 43 shows the Set x Half x Order interaction. It can be seen that the two order conditions show converse patterns of variation. In Order AS, accuracy decreased from Half 1 to Half 2 of blocks of Set A and increased from Half 1 to Half 2 of blocks of Set S. In contrast, in Order SA, accuracy increased from Half 1 to Half 2 of blocks of Set A and decreased from Half 1 to Half 2 of blocks of Set S.

	SET A		SET S	
HALF	1	2	1	2
ORDER AS	94.55	92.85	81.70	84.57
ORDER SA	93.66	96.03	89.62	87.52

TABLE 43. Percentage of stimuli for which responses were correct in each half of blocks of Set A and of Set S in Order AS and in Order SA. The percentages were taken over the 10 observers in each order.

		PERIOD 1		PERIOD 2	
		SET A	SET S	SET A	SET S
ORDER AS					
DISCRIMINABILITY	1	79.38	67.55	83.27	70.90
	2	93.16	80.29	90.94	78.50
	3	99.24	86.61	98.49	81.97
	4	98.22	98.97	97.92	90.07
	5	97.40	92.36	99.03	94.09
ORDER SA					
DISCRIMINABILITY	1	84.52	70.98	79.96	77.48
	2	97.80	82.77	93.64	87.49
	3	98.36	92.34	96.86	91.56
	4	99.33	97.22	99.14	94.99
	5	99.50	96.03	99.36	94.83

TABLE 44. Percentage of stimuli for which responses were correct at each level of discriminability in blocks of Set A and of Set S in Periods 1 and 2 of Order AS and Order SA. The percentages were taken over the 10 observers in each order.

Table 44 shows the Period x Set x Discriminability x Order interaction which reached significance because, in Order AS, accuracy did not show a monotonic increase with stimulus difference in the blocks of Set A in Periods 1 and 2. Similarly, in Order SA, accuracy decreased from level 4 to level 5 of Discriminability in the block of Set S in Period 1.

Non-monotonic changes in accuracy as stimulus difference increased also gave rise to the Period x Set x Half x Discriminability x Response x Order interaction which is shown in table 45. These changes in accuracy did not appear to follow any consistent pattern.

(b) Response time

The Set main effect arose because response times were longer under Set A than under Set S. The mean times for responses made under Sets A and S were 1003.47 msec and 685.10 msec, respectively.

The main effect of Discriminability was significant because response times decreased as stimulus difference increased, but at a decreasing rate. The mean times for responses at Discriminability levels 1 to 5 were 988.91 msec, 887.21 msec, 812.97 msec, 781.28 msec and 751.08 msec, respectively.

The Set x Discriminability interaction is shown in figure 65 where it can be seen that there was no significant change in response time as stimulus difference increased under Set S. In contrast, under Set A, response time decreased significantly from Discriminability levels 1 to 3 but it did not vary from Discriminability levels 3 to 5.

The mean times for right and left responses in Half 1 and Half 2 of blocks in Period 1 and Period 2 are shown in table 46. The Period x Half x Response interaction was significant because times for right responses increased significantly from Half 1 to Half 2 in Period 1 but did not change

			PERIOD 1				PERIOD 2			
			SET A		SET S		SET A		SET S	
HALF			1	2	1	2	1	2	1	2
ORDER AS										
RESPONSE										
RIGHT	DISCRIMINABILITY	1	88.35	82.30	56.41	74.55	88.32	88.10	70.56	68.03
		2	98.07	92.71	72.68	82.11	88.75	92.13	78.49	72.33
		3	100.00	98.50	86.95	94.68	98.45	98.33	84.61	80.85
		4	98.10	100.00	90.03	88.68	100.00	98.33	84.90	96.43
		5	98.81	96.09	90.33	98.00	98.00	100.00	94.59	92.78
LEFT		1	82.56	64.31	68.75	70.48	76.56	80.00	72.81	72.21
		2	88.90	92.97	80.00	86.36	92.30	90.56	74.85	88.33
		3	98.46	100.00	84.17	80.62	100.00	96.67	82.37	80.06
		4	98.23	96.56	86.33	90.83	96.67	96.67	90.06	88.87
		5	100.00	94.68	88.76	92.66	100.00	98.13	96.39	92.61
ORDER SA										
RESPONSE										
RIGHT	DISCRIMINABILITY	1	78.90	80.15	66.83	58.15	70.19	90.33	78.76	68.23
		2	94.54	96.67	84.36	74.33	86.75	90.75	86.59	84.81
		3	98.61	100.00	92.37	88.41	92.59	94.83	88.83	86.38
		4	98.67	100.00	94.50	100.00	98.08	100.00	94.42	92.36
		5	98.00	100.00	98.08	94.44	98.76	100.00	94.25	94.55
LEFT		1	84.38	94.65	76.36	82.57	80.51	78.81	82.41	80.50
		2	100.00	100.00	88.30	84.09	98.69	98.37	92.57	86.00
		3	96.07	98.76	92.14	96.44	100.00	100.00	96.28	94.76
		4	100.00	98.63	100.00	94.38	98.47	100.00	96.36	96.81
		5	100.00	100.00	92.93	98.67	100.00	98.67	96.06	94.44

TABLE 45. Percentage of right and left stimuli for which responses were correct at each level of discriminability in each half of blocks of Set A and of Set S in Periods 1 and 2 of Order AS and Order SA. The percentages were taken over the 10 observers in each order.

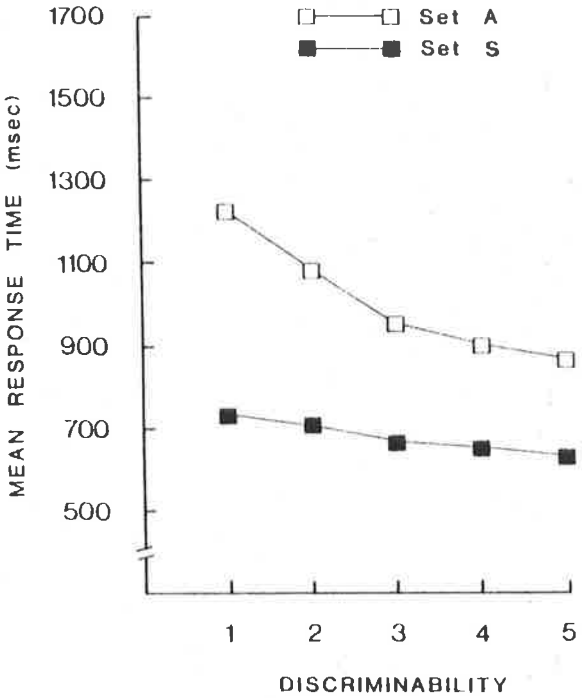


FIGURE 65. Mean times for correct responses at each level of discriminability under Set A (open squares) and Set S (filled squares). The means were taken over 20 observers, the 10 observers in Order AS and the 10 observers in Order SA.

RESPONSE	HALF	PERIOD 1		PERIOD 2	
		1	2	1	2
RIGHT		835.80	871.32	830.07	838.68
LEFT		863.01	837.53	830.28	847.61

TABLE 46. Mean times for correct left and right responses in Half 1 and Half 2 of blocks in Period 1 and Period 2. The means were taken over 20 observers, the 10 observers in Order AS and the 10 observers in Order SA.

		PERIOD 1		PERIOD 2	
RESPONSE		LEFT	RIGHT	LEFT	RIGHT
DISCRIMINABILITY	1	965.81	1008.68	998.04	983.20
	2	891.80	906.27	873.53	877.34
	3	854.82	804.57	781.98	810.50
	4	786.02	788.61	785.88	764.58
	5	752.94	759.80	755.31	736.29

TABLE 47. Mean times for correct left and right responses at each level of discriminability in Period 1 and Period 2. The means were taken over 20 observers, the 10 observers in Order AS and the 10 observers in Order SA.

from Half 1 to Half 2 in Period 2. At the same time, left responses decreased from Half 1 to Half 2 in Period 1 and increased from Half 1 to Half 2 in Period 2.

Table 47 shows the Period x Discriminability x Response interaction. It appears that this interaction resulted from unsystematic changes in response time between the levels of Discriminability. In particular, there was no difference in times for left responses at Discriminability levels 3 and 4 in Period 2. This contrasts with the steady decrease in times for right and left responses in Period 1 and right responses in Period 2 when stimulus difference increased.

Similarly, unsystematic changes in response time between levels of Discriminability in Order AS appear to have given rise to the Half x Discriminability x Response x Order interaction which is shown in table 48.

		ORDER AS				ORDER SA			
	HALF	1		2		1		2	
	RESPONSE	L	R	L	R	L	R	L	R
DISCRIMINABILITY	1	951.68	885.58	893.55	1018.67	1031.14	1047.73	1051.32	1031.58
	2	817.45	835.64	896.89	884.72	928.52	916.25	887.57	930.61
	3	816.22	773.93	799.94	749.04	831.08	845.90	826.37	861.26
	4	788.88	728.15	758.70	773.76	812.18	809.83	794.83	794.66
	5	721.87	709.57	746.62	742.76	767.64	776.85	780.55	763.01

TABLE 48. Mean times for correct left and right responses at each level of discriminability in Half 1 and Half 2 of blocks in Order AS and in Order SA. The means were taken over the 10 observers in each order.

For both right and left responses in both halves of blocks in Order SA, times decreased as stimulus difference increased. In contrast, only times for left responses in Half 2 of blocks in Order AS showed a regular inverse relationship with stimulus difference.

The Set x Half x Response x Order interaction, shown in table 49, arose because times for right and left responses showed apparently irregular increases and decreases from Half 1 to Half 2 of blocks, while in all cases, response times were shorter in Order AS than in Order SA.

Table 50 shows the Set x Half x Discriminability x Response x Order interaction. Once more this interaction arose due to irregular variations in response time between levels of Discriminability. Similarly, the Period x Set x Discriminability x Response x Order interaction (see table 51) and the Period x Set x Half x Discriminability x Order interaction (see table 52) appear to have arisen due to unsystematic changes in response time across levels of Discriminability. There was no clear trend in the pattern of changes for any of these significant interactions.

	SET A				SET S			
HALF	1		2		1		2	
RESPONSE	L	R	L	R	L	R	L	R
ORDER AS	983.46	915.24	982.57	1003.92	654.97	657.90	651.32	663.65
ORDER SA	1031.74	1043.84	1042.25	1024.71	716.48	714.78	693.99	727.73

TABLE 49. Mean times for correct left and right responses in each half of blocks of Set A and Set S performed in Order AS or in Order SA. The means were taken over the 10 observers in each order.

		SET A				SET S			
HALF		1		2		1		2	
RESPONSE		L	R	L	R	L	R	L	R
ORDER AS									
DISCRIMINABILITY	1	1207.38	1097.48	1108.39	1285.34	695.98	673.67	678.71	751.99
	2	991.64	974.14	1097.42	1064.17	643.25	697.13	696.36	705.27
	3	989.90	885.79	953.05	893.88	642.53	662.07	646.07	604.19
	4	915.28	824.52	885.63	917.65	662.48	631.77	610.19	629.87
	5	813.11	794.26	868.35	858.58	630.63	624.88	624.52	626.93
ORDER SA									
DISCRIMINABILITY	1	1252.05	1317.25	1348.48	1224.88	810.22	778.21	754.16	838.28
	2	1120.29	1096.15	1069.04	1088.89	736.75	736.34	706.09	772.32
	3	952.79	1002.11	961.09	1018.28	709.36	689.69	691.64	704.24
	4	942.36	920.83	926.87	909.70	681.99	698.82	662.78	679.61
	5	891.19	882.87	905.77	881.80	644.08	670.82	655.32	644.21

TABLE 50. Mean times for correct left and right responses at each level of discriminability in Half 1 and Half 2 of blocks of Set A and of Set S in Order AS and Order SA. The means were taken over the 10 observers in each order.

		PERIOD 1				PERIOD 2			
		SET A		SET S		SET A		SET S	
RESPONSE		L	R	L	R	L	R	L	R
ORDER AS									
DISCRIMINABILITY	1	1061.44	1167.14	722.55	677.41	1254.33	1215.69	652.15	748.26
	2	1039.42	1003.34	694.44	718.06	1049.64	1034.98	645.17	684.34
	3	1029.47	873.70	645.03	607.83	913.48	905.97	644.34	658.43
	4	902.74	865.68	618.33	640.32	898.16	876.49	654.34	621.32
	5	829.14	811.92	604.87	620.95	852.33	840.91	650.27	630.86
ORDER SA									
DISCRIMINABILITY	1	1292.03	1293.76	787.20	896.01	1308.50	1248.37	777.18	720.48
	2	1101.93	1114.62	731.39	789.04	1087.04	1070.42	712.25	719.62
	3	1002.71	1019.93	742.06	716.82	911.17	1000.46	658.94	677.12
	4	924.86	929.92	698.16	718.53	944.39	900.61	646.61	659.89
	5	903.48	913.99	674.24	692.34	893.48	850.68	625.17	622.69

TABLE 51. Mean times for correct left and right responses at each level of discriminability in blocks of Set A and Set S in Period 1 and Period 2 of Order AS and of Order SA. The means were taken over the 10 observers in each order.

		PERIOD 1				PERIOD 2			
		SET A		SET S		SET A		SET S	
HALF		1	2	1	2	1	2	1	2
ORDER AS									
DISCRIMINABILITY	1	1092.44	1136.18	705.15	694.81	1212.43	1257.59	664.51	735.90
	2	1000.75	1042.01	695.15	717.35	965.04	1119.58	645.23	684.27
	3	1012.01	891.16	617.47	635.39	863.68	955.78	687.13	615.64
	4	878.79	889.63	616.83	641.82	861.01	913.65	677.42	598.25
	5	812.92	828.14	608.64	617.18	794.45	898.79	646.86	634.27
ORDER SA									
DISCRIMINABILITY	1	1323.79	1262.00	820.49	862.72	1245.52	1311.36	767.94	729.72
	2	1117.63	1098.92	736.12	784.31	1098.45	1059.01	736.97	694.90
	3	993.93	1028.70	720.73	738.14	960.96	950.67	678.33	657.73
	4	922.61	932.15	699.89	716.80	940.58	904.42	680.91	625.59
	5	928.45	889.02	684.35	682.23	845.61	898.55	630.55	617.30

TABLE 52. Mean times for correct responses at each level of discriminability in each half of blocks of Set A and of Set S in Period 1 and Period 2. Order AS and Order SA are shown separately. The means were taken over the 10 observers in each order.

Since only 6% of all responses made when performing under Set A were incorrect, data on response times for errors were available for all observers only at the most difficult level of Discriminability, level 1. For this level, error response times under Set A were longer than times for the corresponding correct responses, while error response times under Set S were equal to correct response times for this set. The mean times for correct and incorrect responses were 1003.47 msec and 1317.47 msec, respectively for Set A, and 685.10 msec and 734.15 msec, respectively for Set S. A two-way analysis of variance with repeated measures on both factors showed significant main effects due to the response being correct or incorrect ($F(1,19) = 5.45$, MSerror = 31056, $p < .001$) and the set for speed or accuracy ($F(1,19) = 29.35$, MSerror = 237977, $p < .001$). The interaction of these two factors was also significant ($F(1,19) = 5.02$, MSerror = 33260, $p < .05$).

(c) Confidence

The Set main effect arose because confidence was higher under Set A than under Set S. The mean confidence in Sets A and S was 3.40 and 3.17, respectively.

The Discriminability main effect reached significance because confidence increased with stimulus difference, but at a decreasing rate. The mean confidence at Discriminability levels 1 to 5 was 2.71, 3.04, 3.38, 3.58 and 3.72, respectively.

The Set x Discriminability interaction, shown in figure 66, arose because, at Discriminability levels 3, 4 and 5, confidence in Set A was significantly higher than confidence in Set S. There was no difference in confidence in Sets A and S at Discriminability levels 1 and 2.

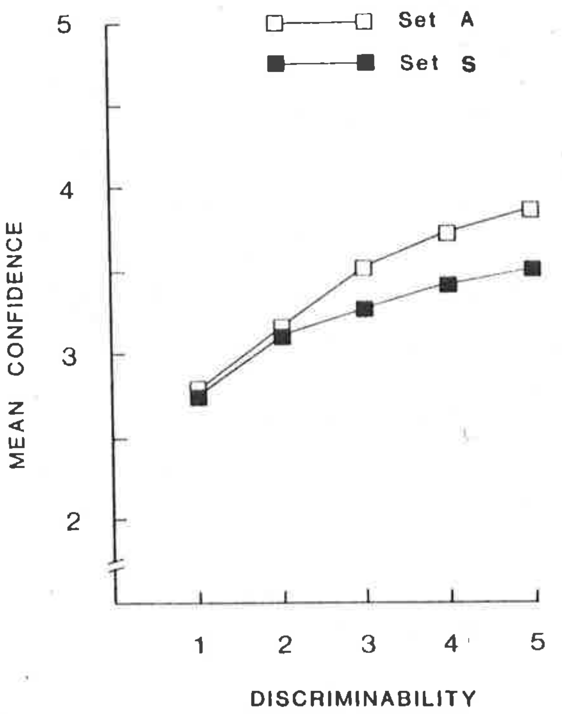


FIGURE 66. Mean confidence in correct responses at each level of discriminability in Set A (open squares) and Set S (filled squares). The means were taken over 20 observers, the 10 observers in Order AS and the 10 observers in Order SA.

The Period x Set x Discriminability interaction, shown in table 53, was significant because, for Set A, confidence at Discriminability levels 1 and 2 decreased from Period 1 to Period 2. In contrast, for Set S, confidence at Discriminability levels 1 and 2 increased from Period 1 to Period 2. Confidence at the other levels of Discriminability did not change from Period 1 to Period 2 for either set.

The interaction of Period x Half x Discriminability x Response (see table 54) resulted from unsystematic changes in confidence during the experimental session. Variations in confidence for both left and right responses from Half 1 to Half 2 of blocks were most marked at Discriminability levels 1 and 2, and did not appear to follow a clear pattern.

	PERIOD 1		PERIOD 2	
DISCRIMINABILITY	SET A	SET S	SET A	SET S
1	2.78	2.56	2.66	2.83
2	3.19	2.86	3.08	3.02
3	3.53	3.28	3.49	3.21
4	3.70	3.47	3.73	3.39
5	3.91	3.57	3.89	3.49

TABLE 53. Mean confidence in correct responses at each level of discriminability in blocks of Set A and Set S in each period. The means were taken over 20 observers, the 10 observers in Order AS and the 10 observers in Order SA.

	PERIOD 1				PERIOD 2			
HALF	1		2		1		2	
RESPONSE	L	R	L	R	L	R	L	R
DISCRIMINABILITY 1	2.58	2.90	2.63	2.58	2.69	2.79	2.60	2.89
2	3.07	3.08	2.99	2.97	2.93	3.10	3.08	3.09
3	3.30	3.42	3.37	3.55	3.29	3.37	3.32	3.43
4	3.52	3.60	3.63	3.58	3.50	3.59	3.44	3.71
5	3.62	3.90	3.61	3.85	3.71	3.73	3.51	3.81

TABLE 54. Mean confidence in correct left and right responses at each level of discriminability in Half 1 and Half 2 of blocks in Period 1 and Period 2. The means were taken over 20 observers, the 10 observers in Order AS and the 10 observers in Order SA.

Similarly, the Period x Set x Half x Response x Order interaction appeared to arise from irregular changes in confidence during the experimental session (see table 55). There was no systematic variation in confidence even within order conditions. Such irregularities in confidence during the experimental session, coupled with non-monotonic changes in confidence as stimulus difference increased, also appeared to give rise to the Period x Set x Half x Discriminability x Response x Order interaction (see table 56).

	PERIOD 1				PERIOD 2			
	SET A		SET S		SET A		SET S	
HALF	1	2	1	2	1	2	1	2
ORDER AS								
RIGHT	3.43	3.15	3.12	3.05	3.20	3.39	3.05	2.98
LEFT	3.21	3.15	2.94	2.86	3.12	3.01	2.82	2.79
ORDER SA								
RIGHT	3.60	3.68	3.36	3.34	3.61	3.56	3.40	3.61
LEFT	3.58	3.57	3.14	3.38	3.55	3.51	3.40	3.43

TABLE 55. Mean confidence in correct right and left responses in Half 1 and Half 2 of blocks of Set A and Set S in each period. Order AS and Order SA are shown separately. The means were taken over the 10 observers in each order.

			PERIOD 1				PERIOD 2			
			SET A		SET S		SET A		SET S	
HALF			1	2	1	2	1	2	1	2
ORDER AS										
RESPONSE										
RIGHT	DISCRIMINABILITY	1	2.96	2.17	2.66	2.59	2.62	2.65	2.50	2.54
		2	3.36	2.91	2.84	2.66	2.86	3.08	2.86	2.69
		3	3.38	3.42	3.29	3.27	3.32	3.53	3.20	2.96
		4	3.70	3.42	3.32	3.29	3.62	3.75	3.29	3.43
		5	3.76	3.81	3.50	3.42	3.59	3.92	3.42	3.28
LEFT		1	2.36	2.52	2.31	2.06	2.36	2.13	2.27	2.54
		2	3.18	2.87	2.81	2.55	2.71	2.68	2.66	2.78
		3	3.41	3.22	3.04	3.11	3.30	3.32	2.91	2.88
		4	3.46	3.53	3.30	3.40	3.53	3.34	3.09	2.82
		5	3.64	3.60	3.22	3.20	3.70	3.60	3.17	2.95
ORDER SA										
RESPONSE										
RIGHT	DISCRIMINABILITY	1	3.13	3.15	2.85	2.40	2.88	2.95	3.17	3.43
		2	3.21	3.30	2.89	3.01	3.36	3.31	3.33	3.27
		3	3.54	3.94	3.46	3.55	3.65	3.61	3.32	3.61
		4	3.79	3.90	3.59	3.68	4.07	3.76	3.38	3.90
		5	4.34	4.10	4.01	4.06	4.10	4.18	3.81	3.86
LEFT		1	3.01	2.91	2.62	3.01	2.93	2.77	3.19	2.96
		2	3.50	3.18	2.79	3.34	3.12	3.50	3.21	3.35
		3	3.57	3.79	3.19	3.35	3.62	3.60	3.33	3.46
		4	3.90	3.87	3.42	3.72	3.91	3.82	3.45	3.79
		5	3.94	4.12	3.67	3.50	4.16	3.88	3.80	3.61

TABLE 56. Mean confidence in correct right and left responses at each level of discriminability in Half 1 and Half 2 of blocks of Set A and Set S in each period. Order AS and Order SA are shown separately. The means were taken over the 10 observers in each order.

Confidence ratings for incorrect responses were available for all observers only at Discriminability level 1. For this level, the mean confidence for all observers for correct and incorrect responses was 3.40 and 2.49, respectively under Set A, and 3.17 and 2.24, respectively under Set S. A two-way analysis of variance with repeated measures on both factors showed a significant effect due to the correctness or incorrectness of responses ($F(1,19) = 24.09$, $MSerror = 0.14$, $p < .01$) with confidence in errors being lower than confidence in correct responses. The influence of Set and the interaction between Correctness and Set were not significant.

(d) Confidence and response time

For each block, confidence ratings were available at each level of discriminability for each observer. These ratings were plotted against the corresponding response time values and straight lines were fitted. Thus, for each of the 20 observers there were five plots, one at each level of discriminability, in each of four blocks of trials, making 400 plots in all. It was found that 268 of the slopes of the best fitting straight lines were negative, showing that there was a predominantly inverse relationship between confidence and response time within blocks ($p < .001$, binomial test). When the 200 plots for each of Set A and Set S were considered separately, the numbers of negative slopes were 154 and 114, respectively ($p < .001$, $p < .05$, respectively, binomial test). Thus, irrespective of set, there was an inverse relationship between confidence and response time within blocks. When individual levels of discriminability were considered separately, combining sets at each level, a significantly higher number of the 20 slopes for each level was negative for four of the five levels of discriminability (see table 57).

DISCRIMINABILITY	NO. OF NEGATIVE SLOPES	p
1	16	< .01
2	17	< .001
3	16	< .01
4	10	n.s.
5	15	< .05

TABLE 57. The number of negative slopes when, for each observer, confidence was plotted against the corresponding response time values for each level of discriminability and straight lines were fitted. Data from the four blocks of experimental trials were considered together. Therefore, there were 20 plots at each level of discriminability.

When discriminability levels were considered for Set A and for Set S separately, the relationship was no longer significant. However, there appeared to be a consistent variation in this relationship between the two sets. The coefficients of slope and the intercepts of the straight lines fitted to the data for Set A and for Set S are shown in table 58. The slopes for Set A were steeper than those for Set S at all except Discriminability level 4, but this was not significant ($p > .05$, binomial test). However, the intercepts from Set A were larger at each level of Discriminability than the corresponding intercepts for Set S ($p < .05$, binomial test).

When the mean confidence for all responses (correct and incorrect) at each level of discriminability was plotted against the corresponding mean times for these responses, the inverse relationship persisted (see figure 67). The relationship varied between the two sets with longer mean response times and higher mean confidence being obtained under Set A.

DISCRIMINABILITY	SLOPE		INTERCEPT	
	SET A	SET S	SET A	SET S
1	-.0013	-.0009	4.304	3.306
2	-.0011	-.0006	4.325	3.446
3	-.0008	-.0006	4.419	3.736
4	-.0004	.0011	4.656	3.178
5	-.0010	-.0004	4.082	3.577

TABLE 58. The coefficients of slope and y-intercepts of straight lines fitted to plots of confidence in correct and incorrect responses against the corresponding response time values at each level of discriminability in each set. For each observer, data from the two blocks performed under each set were considered together and then means were taken over the 20 observers.

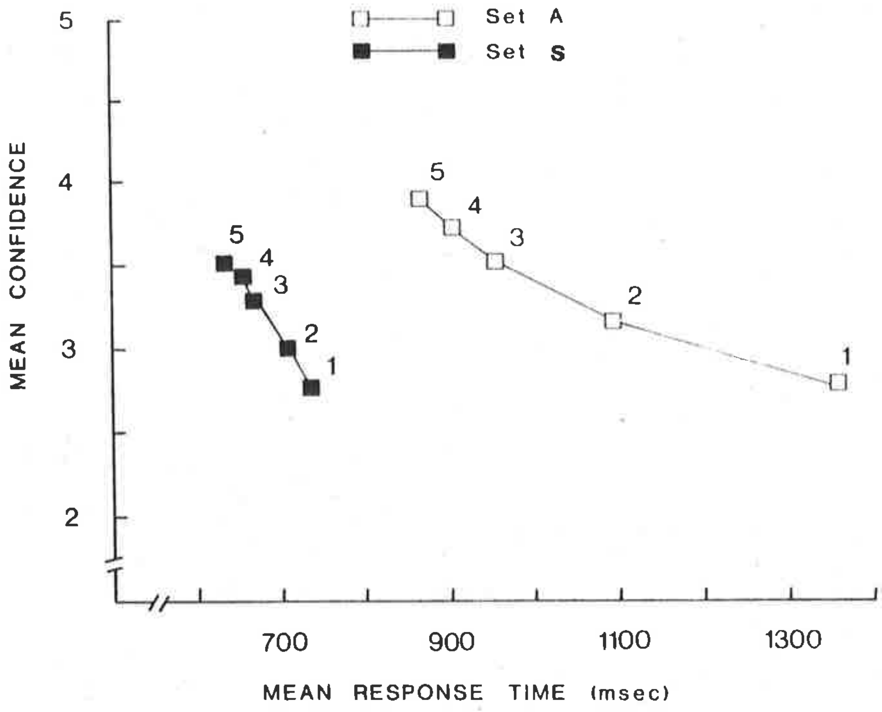


FIGURE 67. Mean confidence for all responses (correct and incorrect) at each level of discriminability under Set A (open squares) and Set S (filled squares) plotted as a function of the corresponding mean response time values. The means were taken over 20 observers, the 10 observers in Order AS and the 10 observers in Order SA.

4. Discussion

The majority of the effects which reached significance did not appear to reflect the manipulation of overall caution by variations in instructions. Instead, these findings appeared to result from unsystematic changes in the response measures between levels of discriminability in the two order conditions. Therefore, although these findings may not be inconsistent with any of the decision processes, they do not correspond to simple predictions of any of them. As a consequence, they will not be discussed in detail. In particular, in view of the difficulty of interpretation of the higher order interactions and since these findings do not follow simply from the experimental manipulation, only main effects and 2-way and 3-way interactions will be discussed.

All three response measures showed evidence of the effects of instructions in a significant Set main effect and a Set x Discriminability interaction. A comparison of the response measures from the two instructional sets showed that the effects of variations in overall caution are consistent with the predictions of the target confidence mechanism. As in the studies of Garrett (1922), Johnson (1939) and Festinger (1943a), accuracy was higher and response times were longer under Set A than under Set S. In contrast to these previous studies, confidence varied with instructions and was higher under Set A than under Set S. This finding lends support to the suggestion of Ascher (1974) and Vickers (1979) that the lack of variation in confidence in previous studies was a product of the form of design used.

The Set x Discriminability interaction arose because, for all three response measures, the change with stimulus difference occurred at a faster rate under Set A than under Set S. This finding is consistent with the results from Wilding's (1974) study and this pattern of variation is evident in simulations of the target confidence mechanism applied to the accumulator model (Vickers, 1979, fig. 77).

In addition, the Discriminability main effect was significant for all of the response measures. The finding that accuracy and confidence increased and response times decreased as stimulus difference increased is a basic finding in discrimination studies, consistent with all of the decision processes, and has been demonstrated in previous experiments in the thesis.

(a) Accuracy

The Set x Discriminability x Response interaction appears to indicate that the effects of instructions were strongest for right responses at the two smallest values of stimulus difference. Right responses were more accurate than left responses at Discriminability level 1 in Set A and right responses were less accurate than left responses at Discriminability levels 1 and 2 in Set S. This finding is consistent with the suggestion made in previous studies in the thesis, that right-handed observers may have a pre-existing preference for right responses which is more apparent at difficult levels of discriminability. Therefore, although this finding does not correspond to a simple prediction of the decision processes, it is not inconsistent with any of them. However, this form of explanation cannot account for the lower accuracy in right responses than in left responses at Discriminability level 2 in Set A. As a consequence, it may be that this effect merely represents unsystematic variations in accuracy which occur when discriminability is low, and as such, it is not related to the experimental manipulation of instructions.

The Set x Half x Order interaction arose because, for both order conditions, accuracy decreased from Half 1 to Half 2 of blocks of the set performed first and increased from Half 1 to Half 2 of blocks of the set performed second. This effect appears to arise from the order in which the blocks were performed rather than the specific instructions given. This would be consistent with an explanation in terms of the target confidence mechanism applied to the accumulator model, according to which the state of the adaptive process at the time of a change in experimental conditions will determine the nature of any changes within conditions. Since all observers performed practice blocks in the same order, it is possible that, for all observers, the basic adaptive process was in the same state at the start of the experimental trials. As a result, it is possible that the variations within blocks will be determined by their position in the series of experimental blocks and will be independent of the specific instructional set which applied in each block. Therefore, all observers will show the same changes within blocks 1, 2, 3 and 4 of the experimental session irrespective of variations in set.

(b) Response time

The Period x Half x Response interaction does not appear to be related to the manipulation of instructions. Instead, it appears to result from irregular changes in times for right and left responses during the progress of the experimental session. Similarly, the Period x Discriminability x Response interaction appears to arise from non-monotonic changes in times for left responses in Period 2 as discriminability increased. This effect does not appear to be related to the manipulation of overall caution by variations in instructions.

When times for correct and incorrect responses are compared, the finding that they did not differ under Set S, but that errors were longer than correct responses under Set A, is consistent with the target confidence mechanism (Vickers, 1979). For example, in Set A when criterion values are high, it would be expected that an error response will be made only after a considerable amount of evidence for both alternatives has been collected, with the result that times will be long. Alternatively, when a correct response is made, it is likely that less evidence in favour of the incorrect alternative will have been collected, so times will be relatively short. Under Set S, when criterion values are low, the collection of evidence for the incorrect alternative will make relatively little difference to the time for the resulting response, whether correct or incorrect.

(c) Confidence

The Period x Set x Discriminability interaction appears to arise from unsystematic variations in confidence at small values of stimulus difference and does not appear to be related to the manipulation of overall caution by instructions.

As for response time data, the finding that confidence was higher in correct responses is consistent with the target confidence mechanism. As described above, when correct responses are made it is likely that little evidence favouring the incorrect alternative will have been collected, so confidence will tend to be high. On the other hand, an error will tend to be made when the amounts of evidence favouring the alternatives are very similar, so confidence will be low.

Although confidence shows a direct relationship with response time between the two instructional sets, this is not the case when the relationship is considered within each set. Within blocks, for four out of five of the levels of discriminability, confidence was inversely related to response time. This would appear to suggest that any formulation for confidence must be able to account for both types of relationship. This suggestion receives further support from the finding of an inverse relationship between confidence and response time across the different levels of discriminability. In addition, since the functions are different for the two sets, it appears that confidence cannot be simply determined by the number of observations taken.

5. Summary and Conclusions

The use of instructions to manipulate overall caution according to a within-subjects design produced clear variations in accuracy, response time and confidence which were consistent with the target confidence mechanism applied to the random walk or accumulator models. The changes in confidence had not been found in previous studies.

For all response measures the variations across levels of discriminability were more marked when caution was high, as found by Wilding (1974), while the relationship between confidence and response time varied within and between conditions. This result pointed to the need for a formulation for confidence which can explain both direct and inverse relationships with response time.

CHAPTER 9

The experiments in this thesis were carried out in order to investigate the variations in criterion values which mediate changes in performance in response to the manipulation of either stimulus properties or instructions. The empirical findings were examined in relation to the predictions of four criterion regulation mechanisms (the ideal observer hypothesis, the response stabilisation hypothesis, the adaptation-level hypothesis and the target confidence mechanism), applied to each of three decision models (Signal Detection Theory, the random walk model and the accumulator model). In all cases, when the effects of each experimental manipulation were considered individually, they were consistent, at least in terms of Signal Detection Theory, with an explanation in terms of variations in criterion values. However, in some cases, the actual changes in performance were consistent with the predictions of several of the decision processes, while in other cases, they were not consistent with the predictions of any of the decision processes. When the findings from all of the experiments are considered together they appear to have implications for both the criterion regulation mechanism and the associated formulation for confidence.

In different experiments, an attempt was made to manipulate bias by variations in a priori probability in both three-category and two-category tasks and by variations in discriminability in two-category tasks. In addition, relative and overall caution were varied by instructions in two-category tasks. The empirical findings from each of these experimental manipulations will be summarised and examined in terms of the adequacy of the theoretical explanations. In each case, the implications of the findings for the criterion regulation mechanisms and the confidence formulations will be discussed.

Methodological problems which appear important for further examination of these issues will then be considered.

For each experiment, those findings which arose as a direct result of the experimental manipulation as predicted by any of the decision processes, have been assigned a number. In each case, the number of the experiment is followed by a digit which refers to a specific finding within that experiment. The findings have been tabulated in terms of these numbers, according to the criterion regulation mechanisms and decision models with which they agree or disagree. Separate tables are given for each form of experimental manipulation and for three-category tasks and two-category tasks where appropriate. In some cases, although a finding is not predicted by any of the decision processes, it may be accommodated by the adoption of assumptions specific to the individual finding. However, such findings cannot contribute to the confirmation or disconfirmation of the predictions of the processes and are not included in the summary tables. In addition, the effects of variations in discriminability were examined both when bias was varied by changes in a priori probability in three-category and two-category tasks, and when instructions were used to vary relative caution and overall caution. In all cases, the empirical findings which resulted from the variations in discriminability were consistent with all of the decision processes. Since they do not differentiate between the processes, these findings are not tabulated.

A. THE MANIPULATION OF BIAS BY VARIATIONS IN A PRIORI PROBABILITY

1. Three-category tasks

(a) Summary of empirical findings

An attempt was made to manipulate bias by variations in a priori probability in three-category tasks in Experiments 1 and 2. In Experiment 1, discriminability was varied within each condition while in Experiment 2, discriminability was held constant.

In Experiment 1, a priori probability was varied according to a between-subjects design. The accuracy of responses to extreme stimuli decreased and the accuracy of responses to equal stimuli increased as the probability of equal stimuli increased (E1:1). This finding agreed with Vickers' (1975) results and was consistent with Signal Detection Theory, the random walk and clock model and the accumulator model in all of which these differences in accuracy are mediated by changes in the cutoffs or the relative values of the criteria for the three responses. In Signal Detection Theory it would be assumed that, as the probability of equal stimuli increases, the cutoffs move to a position where equality responses are more likely to be made. In the random walk and clock model it would be assumed that t decreases, while in the accumulator model, x_e decreases to be low relative to x_g and x_l. As a result, all of the models also predict an increase in the total proportion of equality responses as the probability of equal stimuli increases. This was also found in these data (E1:2). However, although the general form of the psychometric functions was the same as that predicted by all of the models and found in previous studies (Kellogg, 1931; Vickers, 1975) (E1:3), in contrast to the studies of Urban (1910) and Vickers (1975), there was no consistent variation in MP, IU or h as a priori probability changed. When the probability of equal stimuli was less than, or approximately equal to the probability of extreme stimuli, MP and IU increased with the probability of equal stimuli as predicted by all of the models. However, in contrast to the predictions, when equal stimuli became much more probable than extreme stimuli, MP and IU decreased. It was suggested that this may have arisen because, in this condition, observers' assessments of stimulus probabilities were not as extreme as the actual probabilities. As a result, they may have found that they were making more equality responses than they expected. Therefore, in order to reduce the number of equality responses made, they may have adjusted their criteria in the direction opposite to that predicted on the basis of actual stimulus probabilities. The variation in h did not appear to be related to stimulus probability.

As in Vickers' (1975) study, times for equality responses decreased and times for extreme responses increased as the probability of equal stimuli increased (E1:4). In Signal Detection Theory this can be accommodated by the ad hoc choice of a latency function. In the accumulator model, as for accuracy data, it would be assumed that, when equal stimuli have a high probability, x_e takes a value which is low relative to x_g and x_l. However, this finding is not consistent with a random walk and clock model in which t is fixed, and in which either times for equality responses should all be longer than times for extreme responses or they should all be shorter.

Confidence findings in Experiment 1 were less clear. Confidence in equality responses fell on the same straight line as confidence in extreme responses at different values of stimulus difference. This appears to indicate that confidence in both equality and extreme responses is determined by a single mechanism, as is the case in all of the models. At the same time, confidence in both equality and extreme responses was lower when the probability of equal stimuli was low. There was no consistent variation in the relative confidence in equality and extreme responses as a priori probability was manipulated. As a result, the changes in confidence between conditions were not inversely related to the corresponding changes in response time. Therefore, Audley's (1960) hypothesis, applied to either Signal Detection Theory or the random walk and clock model, cannot account for the confidence data. In addition, and in contrast to these data, the accumulator model with the balance of evidence hypothesis predicts that confidence in equality responses should decrease as x_e decreases when the probability of equal stimuli increases. However, confidence measures were being compared between conditions which were performed by different groups of observers. Therefore, the lack of clarity in these findings may be a result of range effects which would be expected in this type of experimental situation (Poulton, 1973, 1975, 1975). As a consequence, any conclusions about these results must be made with reservation.

In Experiment 2, a priori probability was varied according to a within-subjects design. Responses to equal stimuli were less accurate than responses to extreme stimuli, irrespective of a priori probability. This appeared to indicate that, as in earlier studies (e.g. Hayden, 1906; Angell, 1907), there was a bias against equality responses which was not related to the experimental manipulation. As such, this finding is consistent with all of the decision processes in which it would be assumed that the criterion for equality responses was higher than those for extreme responses, throughout the entire experimental session. However, there appeared to be some overriding effect of the experimental manipulation in one order condition. For this group of observers, the accuracy of responses to equal stimuli increased with the probability of equal stimuli. This result was equivalent to the accuracy findings in Experiment 1 and therefore, it can be explained by all of the models in terms of variations in the relative values of criteria for equality and extreme responses. The finding that this effect arose in only one order condition appeared to be a result of the influence of observers' expectations about stimulus probabilities on criterion values. On the basis of practice trials, it seems likely that observers would expect that the stimuli would be equiprobable. Therefore, in order to reduce the discrepancy between these expectations and responding, the group of observers for whom the first experimental block had few equal stimuli, may have adopted a bias towards equality responses. Since this form of compensatory adjustment of criteria works against the overall bias against equality responses, in this group, this form of criterion adjustment may have resulted in the influence of a priori probability becoming apparent. Conversely, in order to reduce the difference between expectations and responding, the group of observers for whom the first experimental block had many equal stimuli may have adopted a bias against equality responses. In this case, the overall bias against equality responses is reinforced and therefore, it may completely mask the effects of a priori probability.

Response time data also showed evidence of an overall bias against equality responses which were made more slowly than extreme responses, irrespective of a priori probability. This finding is consistent with accuracy data and can be accommodated by the models as described above for the accuracy finding. In addition, there was a change in the degree of the bias against equality responses as a priori probability was varied. There was a decrease in the times for equality responses relative to the times for extreme responses when the probability of equal stimuli increased (E2:1). As in

Experiment 1, this can be accommodated in Signal Detection Theory by the ad hoc choice of a latency function and is predicted by the accumulator model, in which there would be a decrease in x_e relative to x_g and x_l. However, as in Experiment 1, these data are not consistent with the random walk and clock model according to which, either times for equality responses should all be longer than times for extreme responses, or they should all be shorter.

Confidence data also showed evidence of an overall bias against equality responses which were less confident than extreme responses, irrespective of a priori probability. This finding was consistent with both accuracy and response time data and can be accommodated by all of the models as described above for the accuracy finding. In addition, confidence varied both within and between blocks of different a priori probability. In general, confidence increased when stimulus probability increased (E2:2), and continued to increase throughout the subsequent block. Since there was no corresponding finding in response time data, Audley's (1960) hypothesis, applied to either Signal Detection Theory or the random walk and clock model, cannot account for this finding. Similarly, this finding is not consistent with the accumulator model, which predicts that the criterion for responses to a stimulus of high probability should be low and therefore, that confidence should be low.

In addition, when response measures were considered within conditions, there was a predominantly inverse relationship between confidence and response time for each of the three responses in each condition. This finding is consistent with Audley's (1960) hypothesis applied to either Signal Detection Theory or the random walk and clock model. Similarly, it is predicted by an accumulator model in which a short response time indicates that the criterion for a given response has been satisfied after only a few observations.

In this case it is likely that the majority of observations favoured that response. Therefore, the difference in the accumulated totals will tend to be large and confidence will be high.

(b) Implications for the decision processes

The findings from Experiments 1 and 2 are summarised in table 59. From the table it appears that the accuracy findings were consistent with variations in the relative values of the criteria for the three responses in each decision process. The response time findings were as predicted by the accumulator model. They could also be accommodated by Signal Detection Theory by the *ad hoc* choice of a latency function. However, in this form, the model merely gives a description of any response time data, and can provide little insight into the nature of the underlying decision process. Meanwhile, the response time data cannot be accommodated by the random walk and clock model, since the inclusion of a response time deadline gives rise to the prediction that all times for equality responses will be longer than times for extreme responses, or all will be shorter, and this prediction receives no support from these data.

From table 59 it appears that confidence data provide little evidence towards an evaluation of the decision processes. In Experiment 1, confidence in equality responses lay on the same straight line as confidence in extreme responses at different values of stimulus difference. This suggests that confidence in both equality and extreme responses is determined by a single mechanism, as is the case in all of the models. However, contrary to Audley's (1960) hypothesis, when measures were taken between conditions in both experiments, confidence did not vary as an inverse function of response time. At the same time, in contrast to the predictions of the accumulator model, confidence increased with stimulus probability in

DECISION PROCESSES	EMPIRICAL FINDINGS		
	ACCURACY	RESPONSE TIME	CONFIDENCE
SIGNAL DETECTION THEORY	E1:1 accuracy increased with _a priori_ probability E1:2 total proportion of equality responses increased with _a priori_ probability of equal stimuli. E1:3 shape of psychometric functions	E1:4 E2:1 response time decreased as _a priori_ probability increased	
RANDOM WALK AND CLOCK MODEL	E1:1 E1:2 E1:3		
ACCUMULATOR MODEL	E1:1 E1:2 E1:3	E1:4 E2:1	[E2:2] confidence increased with _a priori_ probability

TABLE 59: A summary of the findings from Experiments 1 and 2 in which bias was manipulated by variations in _a priori_ probability in three-category tasks. Those findings which arose as a direct result of the experimental manipulation, as predicted by any of the decision processes, were assigned a number in the text. The findings (as represented by these numbers) have been tabulated according to the decision processes with which they are consistent. A brief description of the actual finding is given next to the first entry of each finding. Those findings shown in square brackets are the opposite of the predictions of the decision process against which they appear.

Experiment 2. Finally, when measures were compared within conditions in Experiment 2, there was an inverse relationship between confidence and response time. This finding is consistent with both Audley's (1960) hypothesis and the balance of evidence hypothesis. However, since any formulation for confidence should account for the different relationships with response time within and between conditions, Audley's (1960) hypothesis seems less adequate than the balance of evidence formulation.

Several factors other than the manipulation of a priori probability appear to have influenced the results. For example, in Experiment 2 there was clear evidence of a pre-existing bias against equality responses. As seen in Chapter 1, this finding is consistent with the majority of early studies in which three-category tasks were used, and can be accommodated by all of the models. In addition, it appears that observers can and do vary criterion values in order to respond in accordance with their expectations or assessments of stimulus probabilities, independent of a particular experimental manipulation. The ability of observers to manipulate their criteria in this way resembles the effects induced by changes in instructions in previous studies (e.g. Fernberger, 1914a,b, 1931). An important possibility, suggested by Experiments 1 and 2, is that observers' expectations about stimulus probabilities may have as strong an influence on performance as experimental manipulations of a priori probability.

2. Two-category tasks

(a) Summary of empirical findings

In Experiments 3, 4 and 5, an attempt was made to manipulate bias by variations in a priori probability in two-category tasks. In Experiments 3 and 4, discriminability was varied within each condition, while in Experiment 5, discriminability was held constant.

In Experiment 3, a priori probability was varied according to a between-subjects design. Accuracy was lower for responses to the more probable stimulus (E3:1). This is the opposite of Laming's (1968, 1969) findings and it is not consistent with the ideal observer hypothesis or the target confidence mechanism. According to both of these mechanisms, the criterion for responses to the more probable stimulus should be low, so the accuracy of the corresponding response should be high. However, these data are consistent with the response stabilisation and adaptation-level hypotheses. In terms of the response stabilisation hypothesis, the criterion for responses to the more probable stimulus should be high, while in terms of the adaptation-level hypothesis, the indifference point will move in such a way that the discriminability of the more probable stimulus is reduced.

There was a change in the magnitude of the difference in accuracy as the experimental session progressed. The size of the difference in accuracy of responses to the less and the more probable stimulus increased from block 1 to block 3 and then decreased from block 3 to block 5. This could be explained in terms of the influence of observers' expectations or assessments of the stimulus probabilities on criterion adjustment. In seems likely that, on the basis of practice trials, observers would expect the stimuli to be equiprobable. As a result, criterion values would be very similar in block 1. By block 3, observers could have become aware that they were making more of one response than they expected and, in order to compensate for this apparent inconsistency, they may have adjusted the criterion for responses to the more probable stimulus upward. This would have resulted in lower accuracy for responses to the more probable stimulus. It seems likely that, by block 5, observers would have realised that the stimuli were not equiprobable. Therefore, they would adjust the criterion for responses to the more probable stimulus downward. Given the actual stimulus probabilities, this downward adjustment should continue until the criterion for responses to the more probable stimulus is lower than that for responses to the less probable stimulus. At this point, responses to the more probable stimulus would be more accurate. It appears that this final adjustment had not been completed by block 5.

The finding that responses to the more probable stimulus were faster (E3:2) was not consistent with the explanations for accuracy data. This result is predicted by the ideal observer hypothesis and the target confidence mechanism according to which the criterion for responses to the more probable stimulus is low. However, it cannot be accounted for in terms of the response stabilisation hypothesis in which the criterion for responses to the more probable stimulus is high, or the adaptation-level hypothesis in which the indifference point moves so that the discriminability of the more probable stimulus is reduced.

In addition, confidence was lower for responses to the more probable stimulus (E3:3). As for response time data, this was not consistent with the explanations for the accuracy findings. As it stands, this finding is consistent with the ideal observer hypothesis or the target confidence mechanism if confidence is determined according to the balance of evidence hypothesis. In these mechanisms, since the criterion for responses to the more probable stimulus is low, the maximum possible confidence in these responses is equal to the low criterion value, so mean confidence is likely to be low. Since confidence is varying as a direct function of response time, these data are not consistent with Audley's (1960) hypothesis in combination with any of the criterion regulation mechanisms.

In Experiment 4, bias was manipulated by variations in a priori probability according to a within-subjects design. All three response measures showed changes both between and within conditions. The change in the response measures across block boundaries tended to be in the opposite direction to the change within blocks, so it did not appear to be a continuation of the change within the previous block. In addition, the changes across block boundaries were steeper than the changes within blocks. Therefore, it appears likely that the change across block boundaries represented the response to the variations in a priori probability. The changes within blocks may represent a subsequent adjustment of the criteria in the opposite direction in order to reduce a mismatch between expectations and responding. Alternatively, the form of the basic decision process may give rise to reversals in performance during the block which are not a direct result of any experimental manipulation.

When the measures in the first half of blocks were considered, accuracy was higher for responses to the more probable stimulus (E4:1). This contrasts with the findings from Experiment 3, but it is consistent with Laming's (1968, 1969) results. The accuracy data are predicted by the ideal observer hypothesis and the target confidence mechanism according to which the criterion for responses to the more probable stimulus is low. However, the ideal observer hypothesis cannot account for the subsequent change in accuracy within the probability condition. It is possible that this change arose because the observers' assessments of the stimulus probabilities did not change as rapidly as the actual probabilities. Immediately after a step change in a priori probability, responding may have matched the stimulus probabilities. However, as the block progressed, observers would have become aware that they were making more of one response and fewer of the alternative response than they expected on the basis of their less extreme assessment of the stimulus probabilities. In order to reduce this discrepancy, observers may have adjusted the criteria in the opposite direction, thus giving rise to the variation found in these data. This form of criterion change is not inconsistent with the ideal observer hypothesis. Alternatively, depending on the state of adaptation of the basic decision mechanism of the accumulator model, a step change in a priori probability may give rise to a reversal in performance during the subsequent block.

The response time measures in the first half of blocks were consistent with Experiment 3 and with Laming's (1968, 1969) findings. Times for response to the more probable stimulus were short (E4:2). Therefore, as described above, these data are consistent with the ideal observer hypothesis and the target confidence mechanism, but contrast with the predictions of the response stabilisation and adaptation-level hypotheses. However, once more, the subsequent change during the block is not a simple prediction of the ideal observer hypothesis. As described above, it may correspond to some form of compensatory criterion adjustment in response to an inappropriate assessment of stimulus probabilities. On the other hand, as noted above, the target confidence mechanism applied to the accumulator model predicts cyclical changes in performance which need not be directly related to the experimental manipulation.

In the first half of blocks, confidence tended to be higher for responses to the more probable stimulus. This finding contrasts with the ideal observer hypothesis and the target confidence mechanism when confidence is determined according to the balance of evidence hypothesis. In both of these mechanisms, the criterion for responses to the more probable stimulus will be low, so confidence will be low. However, these data are consistent with the response stabilisation and adaptation-level hypotheses in combination with the balance of evidence hypothesis. According to these hypotheses, the criterion for responses to the more probable stimulus will be high so confidence will tend to be high. The changes in confidence within blocks did not show any regular pattern. In particular, although confidence was inversely related to response time in the first half of blocks, when the entire experimental session was considered, confidence did not vary as an inverse function of response time. Therefore, these data are not consistent with Audley's (1960) hypothesis applied to any of the decision processes.

In Experiment 5, bias was manipulated by variations in a priori probability according to a within-subjects design. In this case, accuracy (E5:1) and confidence (E5:3) increased, and response time decreased (E5:2) when the a priori probability of a stimulus increased. As seen above, both accuracy and response time data are consistent with Laming's (1968, 1969) findings and are predicted by the ideal observer hypothesis or the target confidence mechanism applied to any of the models. These data cannot be explained in terms of the response stabilisation or adaptation-level hypotheses. Confidence data, considered together with accuracy and response time data, are consistent with only the ideal observer hypothesis applied to Signal Detection Theory according to which confidence varies inversely with response time. However, the ideal observer hypothesis is unable to account for the opposite changes in the response measures which occurred within blocks. As described above, these changes may be explained if it is assumed that observers make subsequent adjustments to their criteria in order to account for their expectations.

In each of these experiments, when measures were taken within conditions, there was an inverse relationship between confidence and response time. This finding is consistent with both Audley's (1960) hypothesis and the balance of evidence formulation.

(b) Implications for the decision processes

(i) the criterion regulation mechanisms. The empirical findings from Experiments 3, 4 and 5 are summarised in table 60. From the table it can be seen that, in Experiment 3, accuracy data were consistent with the response stabilisation and adaptation-level hypotheses. In contrast, in Experiments 4 and 5, accuracy data were consistent with the ideal observer hypothesis and the target confidence mechanism. This apparent contradiction may perhaps be explained in terms of a difference in the rate at which observers adapt to the a priori probability values in between-subjects and within-subjects designs. When a priori probability is consistent throughout the experimental session (as in Experiment 3), adaptation appears to be much slower than when a priori probability varies in step changes (as in Experiments 4 and 5). It appears that the step changes in stimulus probability may indicate to observers that a priori probability is being manipulated. As a result, their expectations about the nature of the experimental task may change. Alternatively, it is possible that the use of a between-subjects design in which individual differences are necessarily confounded with the effects of different values of a priori probability may have given rise to this apparently anomalous result. Similarly, range effects, which tend to be stronger in this form of design may be influencing results (Poulton, 1973, 1975, 1979). This seems more likely since any explanation of accuracy data from Experiment 3 in terms of criterion variations contrasts with the response time measures.

From table 60 it can be seen that, for all three experiments, response time findings were consistent with the ideal observer hypothesis and the target confidence mechanism.

TABLE 60:

DECISION PROCESSES		EMPIRICAL FINDINGS			
MODEL	CRITERION REGULATION MECHANISM	ACCURACY	RESPONSE TIME	CONFIDENCE	
				Audley's hypothesis	Balance of evidence hypothesis
SIGNAL DETECTION THEORY	1. ideal observer hypothesis	E4:1 E5:1 accuracy higher for responses to more probable stimulus [E3:1]	E3:2 E4:2 E5:2 responses to more probable stimulus faster	[E3:3] [E5:3]	E3:3 E5:3 confidence lower in responses to more probable stimulus
	2. response stabilisation hypothesis	E3:1 accuracy lower for responses to more probable stimulus [E4:1] [E5:1]	[E3:2] [E4:2] [E5:2]	[E3:3] [E5:3]	[E3:3] [E5:3]
	3. adaptation-level hypothesis	E3:1 [E4:1] [E5:1]	[E3:2] [E4:2] [E5:2]	[E3:3] [E5:3]	[E3:3] [E5:3]
RANDOM WALK MODEL	1. ideal observer hypothesis	E4:1 E5:1 [E3:1]	E3:2 E4:2 E5:2	[E3:3] [E5:3]	E3:3 F5:3
	2. response stabilisation hypothesis	E3:1 [E4:1] [E5:1]	[E3:2] [E4:2] [E5:2]	[E3:3] [E5:3]	[E3:3] [E5:3]
	3. adaptation-level hypothesis	E3:1 [E4:1] [E5:1]	[E3:2] [E4:2] [E5:2]	[E3:3] [E5:3]	[E3:3] [E5:3]
	4. target confidence mechanism	E4:1 E5:1 [E3:1]	E3:2 E4:2 E5:2	[E3:3] [E5:3]	E3:3 E5:3

TABLE 60 (continued)

DECISION PROCESSES		EMPIRICAL FINDINGS			
MODEL	CRITERION REGULATION MECHANISM	ACCURACY	RESPONSE TIME	CONFIDENCE	
				Audley's hypothesis	Balance of evidence hypothesis
ACCUMU-LATOR MODEL	1. ideal observer hypothesis	E4:1 E5:1 [E3:1]	E3:2 E4:2 E5:2	[E3:3] [E5:3]	E3:3 E5:3
	2. response stabilisation hypothesis	E3:1 [E4:1] [E5:1]	[E3:2] [E4:2] [E5:2]	[E3:3] [E5:3]	[E3:3] [E5:3]
	3. adaptation-level hypothesis	E3:1 [E4:1] [E5:1]	[E3:2] [E4:2] [E5:2]	[E3:3] [E5:3]	[E3:3] [E5:3]
	4. target confidence mechanism	E4:1 E5:1 [E3:1]	E3:2 E4:2 E5:2	[E3:3] [E5:3]	E3:3 E5:3

TABLE 60. A summary of findings from Experiments 3, 4 and 5 in which bias was manipulated by variations in *a priori* probability in two-category tasks. Those findings which arose as a direct result of the experimental manipulation, as predicted by any of the decision processes, were assigned a number in the text. The findings (as represented by these numbers) have been tabulated according to the decision processes with which they are consistent. A brief description of the actual finding is given next to the first entry of each finding. Those findings shown in square brackets are the opposite of the predictions of the decision process against which they appear.

Only Experiments 3 and 5 showed systematic changes in confidence as a priori probability was varied. In Experiment 3, confidence data were consistent with either the ideal observer hypothesis or the target confidence mechanism in combination with the balance of evidence hypothesis. In Experiment 5, confidence data were consistent with the ideal observer hypothesis applied to Signal Detection Theory in which case confidence is an inverse function of time.

To this point it would appear that the observer regulates criteria either on the basis of estimates of stimulus probability, as in the ideal observer hypothesis, or attempts to maintain a given level of target confidence. However, the ideal observer hypothesis cannot accommodate the apparent reversal in criterion adjustment which takes place during the block following a step change in a priori probability. As seen above, in terms of the ideal observer hypothesis, the changes which occur across a block boundary should continue throughout the subsequent block. On the other hand, in the accumulator model, the nature of the variation within a block depends on the state of adaptation of the decision mechanism when the step change in a priori probability occurs. For example, simulations of the accumulator model performing an equivalent signal detection task (Vickers, 1981) show that the nature of the variation within blocks may vary from block to block, depending on the parameter specifications. Therefore, although the type of reversal within blocks which is found in simulations of the accumulator model may appear systematic, these measures may have a more restricted significance than it at first appears. As a result, the accumumator model does not necessarily predict the regular reversals during blocks which were observed in Experiment 5.

In Chapter 5 it was suggested that the reversals in response measures during blocks could be explained if the observer was using two independent measures of the experimental situation as the basis for criterion regulation. If these measures varied at different rates they might, in combination, produce a reversal in criterion values. Since a priori probability is being manipulated in these experiments, it seems likely that the observer is keeping some form of tally of the actual probability of the stimuli he has experienced, and it is possible that this is one of the measures used to regulate criteria. There is a measure of actual or estimated stimulus probability in each of the ideal observer, response stabilisation and adaptation-level hypotheses, while an estimate of stimulus probability is indirectly included in the target confidence mechanism due to the influence of average experienced confidence. For example, in terms of the response stabilisation hypothesis this tally would correspond to the measure of cumulative probability, C. This form of measure of the actual stimulus probability will necessarily vary with the same periodicity as the stimulus sequence.

Another measure which appears to influence criterion regulation in the studies in this thesis in which a priori probability was varied but in which the observers were not given any information about its manipulation, is the expectation held by the observer. Several findings in the first five experiments have appeared consistent with the notion that observers regulate criteria in order to reduce a mismatch between their responding and their expectations based on their assessment of the stimulus probability. As noted in the discussion of several different results, the observers' assessments of stimulus probability may be influenced by many factors other than the experimental manipulations. For example, in this experimental situation it appears likely that the stimulus probabilities experienced in practice trials would have a strong influence on the observers' probability estimates. Therefore, there is no necessary relationship between the step changes in a priori probability and the variations in the observers' expectations. As a consequence, it is possible that the combination of a measure of the actual stimulus probabilities with a measure of the observers' assessment of stimulus probabilities in a response stabilisation type of criterion regulation mechanism may produce reversals in criterion values and hence, in response measures within blocks.

For example, as noted above, in the form of design used in Experiment 5, a measure of actual stimulus probability will vary with the same periodicity as a priori probability. However, on the basis of previous experience and the practice trials, it seems likely that observers will expect the stimuli to be equiprobable. Since no information is given about the variations in a priori probability, observers may consider that the apparent inequalities in stimulus probabilities which arise from a step increase in p(R) merely correspond to an unrepresentative subset of trials in a sequence which is equiprobable overall. Therefore, their assessment of the stimulus probabilities may vary during the block on the basis of an overall expectation that the stimuli are equiprobable. For example, when there is a step increase in p(R) observers may continue to expect equiprobable stimuli for several trials. As the block progresses, it should become apparent that these expectations are not confirmed, so the observers may change to expect a majority of right stimuli. However, in the light of an overall expectation that stimuli are equiprobable, it is likely that observers will change again, later in the block, to expect a majority of left stimuli. This expectation may countinue after the subsequent step decrease in p(R). Alternatively, there may be several changes in expectations before the next step change in probability since, at no time will the observer receive confirmation of his expectation of equiprobability.

As a consequence of these fluctuations in expectations, criterion values will be adjusted in different directions within and between blocks, as found in these data.

Alternatively, it is possible that the reversal during blocks is due to the influence of an adaptation-level type of mechanism. In this case, in terms of a random walk type of process, the observer would be adjusting the indifference point as described above, with reference to a measure of all of the observations experienced. However, as seen above, his assessment of the stimulus probabilities also appears to influence criterion regulation. For example, when there are step changes in a priori probability as in Experiment 5, once the observer has become used to the alternations in stimulus probability, it seems likely that he will revise his assessment of stimulus probabilities readily and often. In this situation, the initial response to a step change in a priori probability may be based on this probability estimate. The subsequent reversal during the block may be due to the influence of the observer's adaptation level. It is possible that each of these factors may be acting independently. For example, in terms of a random walk process, the observer's estimate of stimulus probabilities may be controlling the position of the starting point while the adaptation level influences the drift. The same form of mechanism could be incorporated into an accumulator process.

This form of dual mechanism appears to provide a possible explanation for the different findings from between-subjects and within-subjects designs. When a priori probability is varied between subjects, observers are less likely to make regular or large revisions to their probability estimates, while the effect of the adaptation level may be expected to persist throughout the session. Therefore, in this case, the influence of the adaptation level may be stronger than the influence of the probability assessment and so, the criteria may be adjusted in the opposite way to when a within-subjects design is used. This type of explanation would also account for the differences in the findings of Experiments 1 and 3 in both of which, a priori probability was varied between subjects. In Experiment 1, a three-category task was used and accuracy increased with a priori probability, while in Experiment 3 a two-category task was used and accuracy decreased when a priori probability increased. In both cases, it would seem likely that the observer's assessment of stimulus probabilities would not change very much during the experimental session. However, in Experiment 1, it would also seem likely that the adaptation level would remain fairly stable during the session because the two extreme stimuli are always equiprobable. In contrast, since the probabilities of the two stimuli in Experiment 3 differ, the adaptation level will continue to change in the direction of the more probable stimulus throughout the session.

(ii) the confidence formulations. If the observer's assessments of stimulus probability and the adaptation level both play a role in the decision process, then they will necessarily influence all of the response measures, including confidence. However, in addition to this direct influence as the result of the choice of a criterion value, intuitively it seems likely that observers may adjust confidence, perhaps after the decision is made, according to whether the decision confirmed expectations. For example, it seems possible that, irrespective of the evidence on which the decision is based, the observer may reduce the confidence he reports in a decision which does not confirm expectations, or increase the reported confidence in a decision which confirms expectations.

Alternatively, it is possible that the rapid variations which are likely to occur in the observer's probability estimates may give rise to variations in the observer's relative or overall caution. It seems likely that if the observer makes frequent adjustments to his probability assessment for a given response, the relative caution with which that response is made may increase and vice versa. In the form of random walk or accumulator processes suggested above, it is possible that, while the observer's probability assessment is represented by the starting point and the adaptation level is represented by the rate of drift, the relative or overall caution with which he is responding may influence the boundary values. In this form of process it is possible that each of these mechanisms may be more or less influential under different experimental manipulations and that each contributes to the determination of confidence, as suggested by Ascher (1974). Therefore, it is possible that the lack of clarity in confidence data may arise from the conflicting influence of the different factors.

The notion that there are several different types of mechanisms determining performance appears more consistent with the different relationships between confidence and response time which were observed within and between conditions. Confidence was not always an inverse function of response time. In general, the changes in confidence and response time between conditions were inversely related, but the changes in confidence which took place from the first to the second half of a block were not always an inverse function of response time. On the other hand, when confidence and response time measures from within a given condition were examined, there was an inverse relationship. This would appear to indicate that Audley's (1960) hypothesis cannot accommodate the complexity of these data. Instead, it appears that confidence reflects more aspects of the decision process than are incorporated in either of the confidence formulations. It appears that both the time taken to make a response and the evidence collected in favour of a given response are instrumental in the determination of confidence, but that other factors such as the observer's probability estimate and his level of caution may also be important.

B. THE MANIPULATION OF BIAS BY VARIATIONS IN DISCRIMINABILITY

1. Summary of empirical findings

In Experiments 6 and 7, an attempt was made to manipulate bias by variations in relative discriminability in two-category tasks, according to a within-subjects design.

None of the effects which reached significance in Experiment 6 appeared to have arisen from the influence of variations in discriminability. However, from the accuracy data it appeared that the level of discriminability used was not very difficult, so that the data may merely have reflected a ceiling in performance. Therefore, in Experiment 7 the same design was used but discriminability was reduced.

In Experiment 7 there was no evidence of any bias in accuracy data. However, the manipulation of discriminability gave rise to systematic changes in response time and confidence both within and between conditions. This contrasted with the findings of Curry et al. (1977) who could not demonstrate any systematic changes across observers.

The increase in times when responses changed from being paired with a difficult discrimination to being paired with an easy discrimination (E7:1) is consistent with the adaptation-level hypothesis but not with the target confidence mechanism. According to the adaptation-level hypothesis, the effective discriminability of a stimulus is enhanced when it is paired with a difficult discrimination, so fewer observations will be needed before a response is made. On the other hand, in terms of the target confidence mechanism, the criterion for the response to the stimulus paired with a difficult discrimination will be high relative to the criterion for the response to the stimulus paired with an easy discrimination.

The confidence data were less clear. The only clear change was the decrease in confidence which occurred when the context changed from an easy to a difficult discrimination. However, in contrast to response time findings, this change was opposite to the predictions of both the adaptation-level hypothesis and the target confidence mechanism. It appeared that observers may have modified their confidence before reporting it in order to correspond to their expectations about the relative values of confidence at different levels of discriminability. For example, when a stimulus was paired with a difficult discrimination it seems likely that confidence in the appropriate response would be rated relative to the difficult discrimination, and therefore it may be adjusted upwards. The opposite adjustment might occur when a stimulus was paired with an easy discrimination. Alternatively, it is possible that when a stimulus is paired with a difficult discrimination which will not always be accurately discriminated, the observer perceives this stimulus as less probable. Therefore, in terms of the random walk or accumulator type of process considered above, in this experimental situation, the position of the starting point may vary as well as the drift. As a consequence, the nature of variations in confidence will depend on the frequency and extent of revisions to the observers' probability assessments.

In addition, there were problems accounting for the different patterns of change in response time and confidence within conditions. For response time, the changes within conditions formed a continuation of the changes between conditions as predicted by the adaptation-level hypothesis. However, for confidence data, the changes within conditions varied from block to block. This is not inconsistent with an accumulator type of process in which the state of adaptation of the decision mechanism at the time when discriminability changes will determine the direction of any change during the subsequent block.

When measures were examined within conditions, there was an inverse relationship between confidence and response time. This finding is consistent with Audley's (1960) hypothesis and the balance of evidence hypothesis.

2. Implications for the decision processes

The empirical findings from Experiments 6 and 7 are summarised in table 61. From the table it can be seen that, in general, the findings were consistent with the adaptation-level mechanism. In other words, when bias is manipulated by variations in discriminability it appears that observers regulate criteria with reference to an average of all of the stimulus information they receive from observations. However, it appears that other aspects of the stimulus situation may also be important. For example, the apparently anomalous confidence data could be explained if the observer's assessment of stimulus probabilities was also influencing the decision process. As suggested above, it is possible that the adaptation level is reflected in the indifference point and therefore in the drift of a random walk or an accumulator process, while the observer's assessment of stimulus probabilities is represented by the starting point.

DECISION PROCESSES		EMPIRICAL FINDINGS			
MODEL	CRITERION REGULATION MECHANISM	ACCURACY	RESPONSE TIME	CONFIDENCE	
				Audley's hypothesis	Balance of evidence hypothesis
SIGNAL DETECTION THEORY	1. adaptation-level hypothesis		E7:1 response times longer when paired with easy discrimination		
RANDOM WALK MODEL	1. adaptation-level hypothesis		E7:1		
	2. target confidence mechanism		[E7:1]		
ACCUMU-LATOR MODEL	1. adaptation-level hypothesis		E7:1		
	2. target confidence mechanism		[E7:1]		

TABLE 61: A summary of the findings from Experiments 6 and 7 in which bias was manipulated by variations in discriminability in two-category tasks. Those findings which arose as a direct result of the experimental manipulation, as predicted by any of the decision processes, were assigned a number in the text. The findings (as represented by these numbers) have been tabulated according to the decision processes with which they are consistent. A brief description of the actual finding is given next to the first entry of each finding. Those findings shown in square brackets are the opposite of the predictions of the decision process against which they appear.

Further support for the notion of several factors which each influence the decision process is found in an examination of the variation in response time and in confidence within and between conditions. The changes in response time within conditions formed a continuation of the changes between conditions. Therefore, these measures would appear to be providing a direct indication of the variation in the adaptation level and hence in the indifference point in response to the manipulation of discriminability. In particular, it appears that the change in the adaptation level in response to a change in discriminability continues throughout the subsequent block.

In contrast, the changes in confidence between conditions was opposite to the predictions of either the adaptation-level hypothesis or the target confidence mechanism. As noted above, it seems possible that in this case the response measure indicates the combined variation of both the adaptation level and the observer's assessment of stimulus probabilities. In addition, there was no clear pattern in the changes in confidence within blocks. This may arise from the irregular way in which observers revise their probability assessments in an experimental situation in which a priori probability is not in fact varying, but in which actual response probability changes.

C. THE MANIPULATION OF RELATIVE AND OVERALL CAUTION BY INSTRUCTIONS

1. Relative caution

(a) Summary of empirical findings

In Experiments 8 and 9, relative caution was manipulated by changes in instructions in two-category tasks, according to a within-subjects design. In Experiment 8, discriminability was varied within conditions while, in Experiment 9, discriminability was held constant.

In Experiment 8, the effects of instructions were evident only in response time measures. Observers were faster on those responses that instructions indicated should be made quickly, and slower on those responses that instructions indicated should be made accurately (E8:1). There were no changes in response time within conditions. The response time findings were consistent with the ideal observer hypothesis and the target confidence mechanism in both of which instructions to make a response quickly will give rise to a decrease in the criteria for that response. However, these decision processes also predict concomitant changes in accuracy and confidence.

In Experiment 9, both accuracy and response time showed the influence of instructions and, once more, there was no variation within conditions. Accuracy was low (E9:1) and times were long (E9:2) when instructions stressed the accuracy of a response. These data are consistent with the ideal observer hypothesis and the target confidence mechanism. In both cases, instructions stressing accuracy will increase the criterion for the response. As a result, accuracy will be low and times will be long.

Confidence data did not vary with instructions. It appeared that this may have resulted from the stress placed on speed rather than accuracy in the original instructions. As a consequence, it appears that response times did not vary very much between conditions. Since it would be expected that confidence would be related to response time, this may have resulted in considerable overlap in confidence ratings between conditions. However, as predicted by both of the confidence formulations, there was an inverse relationship between confidence and response time within blocks of trials.

(b) Implications for the decision processes

The empirical findings from Experiments 8, 9 and 10 are summarised in table 62. From the table it can be seen that, when instructions are used to vary relative caution it appears that observers use information about the values or costs of the respective responses or about the required level of confidence in each response to regulate criteria. Therefore, the regulation of criterion values is not dependent on the observer's experience of the stimulus sequence. Instead, criteria may be adjusted to take the appropriate values prior to the first stimulus in a condition. This would appear to be the case in these data since there is no evidence of any changes in the response measures within a condition. This contrasts with previous studies in this thesis and appears to highlight the difference between the manipulation of stimulus properties without telling observers about it and the use of instructions. Indeed, this would appear to support the suggestion that the effect of instructions is to give rise to variations in boundary values in a random walk or accumulator type of process, while probability estimates act on the starting point and the adaptation level determines the drift. In this type of process, each of these adjustments could occur independently, and it appears that, in this case, it need only be supposed that the boundaries vary.

DECISION PROCESSES		EMPIRICAL FINDINGS			
MODEL	CRITERION REGULATION MECHANISM	ACCURACY	RESPONSE TIME	CONFIDENCE	
				Audley's hypothesis	Balance of evidence hypothesis
SIGNAL DETECTION THEORY	1. ideal observer hypothesis		E8:1 E9:2 E10:2 responses slower when accuracy instructions		
RANDOM WALK MODEL	1. ideal observer hypothesis		E8:1 E9:2	[E10:3]	
	2. target confidence mechanism	E9:1 E10:1 accuracy high when instructions stress accuracy	E8:1 E9:2		E10:3 confidence low when accuracy instructions
ACCUMULATOR MODEL	1. ideal observer hypothesis		E8:1 E9:2	[E10:3]	
	2. target confidence mechanism	E9:1 E10:1			E10:3

TABLE 62: A summary of the findings from Experiments 8, 9 and 10 in which relative and overall caution were manipulated by variations in instructions in two-category tasks. Those findings which arose as a direct result of the experimental manipulation, as predicted by any of the decision processes, were assigned a number in the text. The findings (as represented by these numbers) have been tabulated according to the decision processes with which they are consistent. A brief description of the actual finding is given next to the first entry of each finding. Those findings shown in square brackets are the opposite of the predictions of the decision process against which they appear.

2. Overall caution

(a) Summary of empirical findings

In Experiment 10, overall caution was varied by instructions according to a within-subjects design in a two-category task. Discriminability was varied within conditions.

In this case, all three response measures showed consistent changes between conditions and did not vary within conditions. Accuracy was high (E10:1), response times were long (E10:2) and confidence was high (E10:3) when instructions stressed the accuracy of performance. The accuracy and response time findings confirm those from previous studies by Garrett (1922), Johnson (1939) and Festinger (1943a) while the variation in confidence contrasts with the findings from these earlier studies. It appears that the use of a within-subjects design in this type of experiment avoids the problem of between session range effects and therefore, the comparison of confidence between conditions is more powerful. All of the findings are consistent with the target confidence mechanism according to which criteria for both responses will be high when accuracy is emphasised. As predicted by both of the confidence formulations, confidence and response time were inversely related within blocks.

(b) Implications for the decision processes

The empirical findings from Experiment 10 are shown in table 62. In this experimental situation the observer appears to regulate criterion values in accordance with a target confidence value chosen on the basis of instructions. As in the previous experiments in which instructions were varied, response measures did not change within conditions. This is consistent with the random walk or accumulator type of process described above in which variations in caution give rise to variations in the boundaries. As in Experiments 8 and 9, it appears that neither probability estimates nor the adaptation level influence performance in this situation. Therefore, in this case, it seems likely that confidence will be determined on the basis of the position of the boundaries alone. This may account for the clarity of confidence results in this study.

D. CONCLUSIONS AND FUTURE RESEARCH

From the review of the literature presented in Chapters 1 and 2, it appeared that the notion of a variable criterion was the critical element in models developed to account for variations in discrimination and identification performance which resulted from the manipulation of stimulus properties or instructions. It appeared that each of the dependent variables of accuracy, response time and confidence, had a simple relationship to the change in criterion values. For example, in studies of the manipulation of caution, Garrett (1922), Johnson (1939) and Festinger (1943a) found changes in both accuracy and response time which appeared to indicate that the criteria for both responses increased with caution. The confidence measures in these studies did not show a corresponding change but this was interpreted as an indication of the choice of an inappropriate experimental design (Ascher, 1974; Vickers, 1979) rather than that confidence may not be responsive to variations in caution.

In addition, from the literature it appeared that the response measures could be used individually to indicate the nature of variation in criterion values. As seen in Chapter 1, in early studies, accuracy alone was used as an index of performance. More recently, the complex nature of the relationship between accuracy and response time has been recognised and studied extensively (e.g. Philip, 1936, 1947; Vickers et al., 1971; Swensson, 1972a,b; Pachella, 1974). However, from the experiments in the thesis it appears that confidence also has a rather less straightforward relationship to accuracy and response time than was at first thought. For example, the relationship between confidence and response time was different when measured within or between the conditions of a given experimental manipulation and the specific relationships observed both within and between conditions depended on the specific experimental manipulation.

Similarly, in the literature, much attention has been paid to an examination of the adequacy of the decision models and, more recently, the criterion regulation mechanisms. In these studies the aim has appeared to be to arrive at a single self-regulating decision mechanism which can account for all variations in performance in discrimination tasks. Once more, from the experiments presented here this would appear to be an over-simplification of the problem. Indeed, when the experiments in the thesis are considered together, it appears that several different factors from the experimental situation contribute to the decision process.

In those experiments in which a priori probability was varied it appeared that the observer's estimate of stimulus probabilities was a major determinant of any variation in criterion values. Therefore, these data suggested that an ideal observer type of mechanism or a response stabilisation hypothesis might be the most appropriate form of criterion regulation mechanism in this situation. However, the changes in response measures within conditions could not be explained in terms of the operation of any single mechanism. They appeared to be most consistent with the notion that there was more than one mechanism influencing performance. In particular, it appeared that the reversals during blocks were consistent with an adaptation-level type of process. It was suggested that, in terms of a random walk or accumulator type of model it was possible that the effects of the observer's assessment of stimulus probability may influence the position of the starting point, while the value of the adaptation level would determine the indifference point and hence the rate of drift.

When bias was manipulated by changes in discriminability the major determinant of criterion regulation appeared to be the value of the adaptation level. However, once more it appeared that other mechanisms were playing a part in criterion regulation. In this case, confidence ratings appeared to have been influenced by the observer's assessment of stimulus probabilities. As described above, these two factors may both be incorporated into a single decision process.

Finally, when relative and overall caution were varied with instructions the observer appeared to regulate criteria in order to maintain a given value of target confidence which was adopted at the beginning of the session to correspond to instructions. Therefore, it appears that the effects of variations in instructions are different from the influence of changes in either a priori probability or discriminability. In the random walk or accumulator type of process described above, it is possible that the influence of instructions is represented by movements in the boundary values.

In terms of the random walk or accumulator type of process described above, as suggested by Ascher (1974) confidence is a product of variations in the starting point, the drift and the boundaries. In other words, changes in probability estimates, discriminability and instructions or perceived costs and payoffs will all influence confidence. Therefore, since all of these factors appear to operate simultaneously, the time course of the effects of any of these variations is critical to the determination of the exact independent variable which is in operation. As a result, it appears that future research should concentrate on a more detailed examination of the rate at which the response measures change. It is possible that a trial by trial analysis of the response measures would permit the description of several different components in the decision process. At the same time, the measurement of the time to make confidence ratings may elucidate the relationship between the hypothetical basis for confidence and reported confidence.

Therefore, it appears that future research might be most rewarding if it examines performance at a more microscopic level. In contrast, the resulting models of performance must incorporate the variety of possible types of information which appear to affect response measures.

INDEX OF APPENDICES

INDEX OF APPENDICES (continued)

APPENDIX 1 : EXPERIMENT 1. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 1.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	6.62	3	2.21	5.46	< .001
Error	12.92	32	0.40		
Discriminability (D)	54.53	5	10.91	114.28	< .001
C x D	3.86	15	0.26	2.70	< .01
Error	15.27	160	0.10		

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 1.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	240719	3	80240	1.54	
Error	1664297	32	52009		
Discriminability (D)	1071143	5	214229	38.58	< .001
C x D	460660	15	30711	5.53	< .001
Error	888425	160	5553		

C. Summary table of the analysis of variance of mean times for responses to equal stimuli from Experiment 1.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	308931	3	102977	2.51	
Error	1313089	32	41034		
Correctness (Co)	102748	1	102748	18.72	< .01
C x Co	116192	3	38731	7.06	< .01
Error	175629	32	5488		

D. Summary table of the analysis of variance of mean times for equality responses from Experiment 1.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	902701	3	300900	4.93	< .01
Error	1953910	32	61060		
Correctness (Co)	3	1	3	0.01	
C x Co	13536	3	4512	1.19	
Error	121678	32	3802		

APPENDIX 1 (continued)

E. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 1.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	82.91	3	27.84	2.67	
Error	331.08	32	10.35		
Discriminability (D)	201.11	5	40.22	75.99	< .001
C x D	14.97	15	1.00	1.89	< .05
Error	84.69	160	0.53		

F. Summary table of the analysis of variance of mean confidence in responses to equal stimuli from Experiment 1.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	44.28	3	14.76	4.28	< .05
Error	110.35	32	3.45		
Correctness (Co)	0.01	1	0.01	0.04	
C x Co	2.43	3	0.81	2.03	
Error	12.78	32	0.40		

G. Summary table of the analysis of variance of mean confidence in equality responses from Experiment 1.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	33.26	3	11.09	2.93	< .05
Error	121.03	32	3.78		
Correctness (Co)	0.93	1	0.93	5.33	< .05
C x Co	0.18	3	0.06	0.35	
Error	5.60	32	0.18		

APPENDIX 2 : EXPERIMENT 2. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 2.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	0.52	1	0.52	0.84	
Error	11.07	18	0.61		
Period (P)	1.38	1	1.38	10.14	< .001
O x P	0.01	1	0.01		
Error	2.45	18	0.14		
Probability (Pr)	0.09	1	0.09	0.66	
O x Pr	0.01	1	0.01	0.01	
Error	2.36	18	0.13		
Half (H)	0.07	1	0.07	0.57	
O x H	0.01	1	0.01	0.10	
Error	2.23	18	0.12		
Response (R)	17.73	2	8.87	12.39	< .001
O x R	0.73	2	0.37	0.51	
Error	25.76	36	0.72		
P x Pr	0.04	1	0.04	0.83	
O x P x Pr	0.01	1	0.01	0.16	
Error	0.91	18	0.05		
P x H	0.19	1	0.19	2.09	
O x P x H	0.00	1	0.00	0.04	
Error	1.60	18	0.09		
Pr x H	0.06	1	0.06	1.23	
O x Pr x H	0.00	1	0.00	0.02	
Error	0.86	18	0.05		
P x R	0.31	2	0.16	1.74	
O x P x R	0.20	2	0.10	1.10	
Error	3.25	36	0.09		
Pr x R	0.06	2	0.03	0.19	
O x Pr x R	1.56	2	0.78	5.16	< .05
Error	5.43	36	0.15		
H x R	0.45	2	0.22	1.61	
O x H x R	0.01	2	0.01	0.03	
Error	5.02	36	0.14		
P x Pr x H	0.31	1	0.31	3.09	
O x P x Pr x H	0.18	1	0.18	1.87	
Error	1.78	18	0.10		
P x Pr x R	0.21	2	0.11	0.95	
O x P x Pr x R	0.61	2	0.30	2.73	
Error	4.02	36	0.11		
P x H x R	0.55	2	0.28	3.06	
O x P x H x R	0.43	2	0.21	2.36	
Error	3.25	36	0.09		
Pr x H x R	0.36	2	0.18	1.98	
O x Pr x H x R	0.01	2	0.01	0.06	
Error	3.28	36	0.09		
P x Pr x H x R	0.04	2	0.02	0.18	
O x P x Pr x H x R	0.57	2	0.28	2.78	
Error	3.67	36	0.10		

APPENDIX 2 (continued)

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 2.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	965307	1	965307	0.76	
Error	23007094	18	1278172		
Period (P)	52679	1	52679	0.90	
O x P	3689	1	3689	0.06	
Error	1050505	18	58361		
Probability (Pr)	49258	1	49258	1.12	
O x Pr	709	1	709	0.02	
Error	793278	18	44071		
Half (H)	8178	1	8178	1.06	
O x H	10369	1	10369	1.34	
Error	139422	18	7746		
Response (R)	1155357	2	577679	6.63	< .01
O x R	62122	2	31061	0.36	
Error	3138774	36	87188		
P x Pr	11074	1	11074	0.53	
O x P x Pr	29248	1	29248	1.41	
Error	374398	18	20800		
P x H	34665	1	34665	1.27	
O x P x H	4931	1	4931	0.18	
Error	493225	18	27401		
Pr x H	47702	1	47702	1.65	
O x Pr x H	37611	1	37611	1.30	
Error	519486	18	28860		
P x R	38567	2	19284	1.06	
O x P x R	3714	2	1857	0.10	
Error	655251	36	18201		
Pr x R	365013	2	182507	9.91	< .001
O x Pr x R	41492	2	20746	1.13	
Error	663177	36	18422		
H x R	4028	2	2014	0.20	
O x H x R	16005	2	8003	0.79	
Error	366524	36	10181		
P x Pr x H	7179	1	7179	0.77	
O x P x Pr x H	21375	1	21375	2.30	
Error	167224	18	9290		
P x Pr x R	16628	2	8314	0.41	
O x P x Pr x R	22543	2	11272	0.56	
Error	723944	36	20110		
P x H x R	134675	2	67338	4.11	< .05
O x P x H x R	57004	2	28502	1.74	
Error	590147	36	16393		
Pr x H x R	29584	2	14792	0.78	
O x Pr x H x R	2184	2	1092	0.06	
Error	685559	36	19043		
P x Pr x H x R	25181	2	12590	1.03	
O x P x Pr x H x R	6370	2	3185	0.26	
Error	438807	36	12189		

APPENDIX 2 (continued)

C. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 2.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	0.35	1	0.35	0.03	
Error	214.03	18	11.89		
Period (P)	1.72	1	1.72	2.68	
O x P	0.24	1	0.24	0.37	
Error	11.54	18	0.64		
Probability (Pr)	0.05	1	0.05	0.10	
O x Pr	0.05	1	0.05	0.12	
Error	8.09	18	0.45		
Half (H)	0.09	1	0.09	0.50	
O x H	0.05	1	0.05	0.27	
Error	3.40	18	0.19		
Response (R)	9.81	2	4.91	4.70	< .05
O x R	1.32	2	0.66	0.63	
Error	37.62	36	1.05		
P x Pr	0.07	1	0.07	0.26	
O x P x Pr	0.00	1	0.00	0.00	
Error	5.04	18	0.28		
P x H	0.24	1	0.24	1.94	
O x P x H	0.02	1	0.02	0.16	
Error	2.22	18	0.12		
Pr x H	0.16	1	0.16	1.17	
O x Pr x H	0.04	1	0.04	0.29	
Error	2.44	18	0.14		
P x R	0.07	2	0.03	0.23	
O x P x R	0.02	2	0.01	0.05	
Error	5.23	36	0.15		
Pr x R	0.46	2	0.23	1.29	
O x Pr x R	0.03	2	0.01	0.07	
Error	6.37	36	0.18		
H x R	0.00	2	0.00	0.02	
O x H x R	0.06	2	0.03	0.39	
Error	2.89	36	0.08		
P x Pr x H	0.01	1	0.01	0.03	
O x P x Pr x H	0.24	1	0.24	1.67	
Error	2.57	18	0.14		
P x Pr x R	0.30	2	0.15	1.89	
O x P x Pr x R	0.19	2	0.10	1.22	
Error	2.83	36	0.08		
P x H x R	0.10	2	0.05	0.50	
O x P x H x R	0.44	2	0.22	2.24	
Error	3.51	36	0.10		
Pr x H x R	0.82	2	0.41	3.90	< .05
O x Pr x H x R	0.00	2	0.00	0.00	
Error	3.81	36	0.11		
P x Pr x H x R	0.07	2	0.04	0.24	
O x P x Pr x H x R	0.00	2	0.00	0.00	
Error	5.47	36	0.15		

APPENDIX 3 : EXPERIMENT 3. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 3.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	6.61	4	1.65	1.65	
Error	45.22	45	1.00		
Discriminability (D)	214.91	4	53.73	340.15	< .001
C x D	3.13	16	0.20	1.24	
Error	28.43	180	0.16		
Block (B)	2.17	4	0.54	3.53	< .01
C x B	3.44	16	0.22	1.44	
Error	26.83	180	0.15		
D x B	3.34	16	0.21	1.47	
C x D x B	6.34	64	0.10	0.70	
Error	102.51	720	0.14		
Response (R)	4.87	1	4.87	3.21	
C x R	53.14	4	13.28	8.75	< .001
Error	68.31	45	1.52		
R x B	0.87	4	0.22	1.14	
C x R x B	3.93	16	0.25	1.29	
Error	34.34	180	0.19		
R x D	0.19	4	0.05	0.19	
C x R x D	64.84	16	4.05	16.05	< .001
Error	45.46	180	0.25		
R x D x B	3.68	16	0.23	1.61	
C x R x D x B	13.89	64	0.22	1.52	< .01
Error	102.63	720	0.14		

APPENDIX 3 (continued)

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 3.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	4314018	4	1078505	2.99	< .05
Error	16221126	45	360469		
Discriminability (D)	8104175	4	2026044	63.99	< .001
C x D	498328	16	31146	0.98	
Error	5699573	180	31664		
Block (B)	160935	4	40234	1.21	
C x B	462828	16	28927	0.87	
Error	5960949	180	33116		
D x B	463847	16	28990	1.52	
C x D x B	1553355	64	24271	1.28	
Error	13689616	720	19013		
Response (R)	841234	1	841234	3.18	
C x R	4374972	4	1093743	4.14	< .01
Error	11899388	45	264431		
R x B	133337	4	33334	1.60	
C x R x B	432353	16	27022	1.30	
Error	3749524	180	20831		
R x D	246085	4	61521	1.57	
C x R x D	273749	16	17109	0.44	
Error	7072615	180	39292		
R x D x B	384673	16	24042	1.22	
C x R x D x B	1500885	64	23451	1.19	
Error	14199179	720	19721		

APPENDIX 3 (continued)

C. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 3.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Condition (C)	14.13	4	3.53	0.17	
Error	950.43	45	21.12		
Discriminability (D)	275.11	4	68.78	135.26	< .001
C x D	5.19	16	0.32	0.64	
Error	91.53	180	0.51		
Block (B)	11.61	4	2.90	5.37	< .001
C x B	6.43	16	0.40	0.74	
Error	97.26	180	0.54		
D x B	2.16	16	0.14	0.92	
C x D x B	11.57	64	0.18	1.23	
Error	106.14	720	0.15		
Response (R)	0.01	1	0.01	0.00	
C x R	96.39	4	24.10	10.71	< .001
Error	101.21	45	2.25		
R x B	1.92	4	0.48	1.71	
C x R x B	5.90	16	0.37	1.32	
Error	50.49	180	0.28		
R x D	1.97	4	0.49	2.10	
C x R x D	18.86	16	1.18	5.04	< .001
Error	42.09	180	0.23		
R x D x B	3.10	16	0.19	1.51	
C x R x D x B	7.96	64	0.12	0.97	
Error	92.46	720	0.13		

APPENDIX 4 : EXPERIMENT 4. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 4.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	3.76	1	3.76	1.13	
Error	59.66	18	3.31		
Period (P)	0.27	1	0.27	1.17	
P x O	0.21	1	0.21	0.92	
Error	4.08	18	0.23		
Probability (Pr)	0.85	1	0.85	1.35	
Pr x O	0.31	1	0.31	0.50	
Error	11.25	18	0.62		
P x Pr	0.77	1	0.77	3.15	
P x Pr x O	0.40	1	0.40	1.65	
Error	4.38	18	0.24		
Half (H)	0.01	1	0.01	0.01	
H x O	0.00	1	0.00	0.00	
Error	3.11	18	0.17		
P x H	0.40	1	0.40	2.80	
P x H x O	0.30	1	0.30	2.10	
Error	2.57	18	0.14		
Pr x H	0.01	1	0.01	0.01	
Pr x H x O	1.11	1	1.11	2.67	
Error	7.45	18	0.41		
P x Pr x H	0.00	1	0.00	0.00	
P x Pr x H x O	0.67	1	0.67	1.68	
Error	7.20	18	0.40		
Discriminability (D)	105.49	4	26.37	70.12	< .001
D x O	0.74	4	0.19	0.49	
Error	27.08	72	0.38		
P x D	1.30	4	0.33	1.27	
P x D x O	2.27	4	0.57	2.21	
Error	18.49	72	0.26		
Pr x D	0.40	4	0.10	0.25	
Pr x D x O	1.41	4	0.35	0.86	
Error	29.41	72	0.41		
P x Pr x D	0.92	4	0.23	0.88	
P x Pr x D x O	1.66	4	0.42	1.57	
Error	18.96	72	0.26		
H x D	1.10	4	0.28	0.91	
H x D x O	1.00	4	0.25	0.83	
Error	21.78	72	0.30		
P x H x D	0.80	4	0.20	0.81	
P x H x D x O	1.21	4	0.30	1.22	
Error	17.76	72	0.25		
Pr x H x D	0.45	4	0.11	0.46	
Pr x H x D x O	0.70	4	0.17	0.72	
Error	17.51	72	0.24		
P x Pr x H x D	4.38	4	1.09	3.38	< .05
P x Pr x H x D x O	0.52	4	0.13	0.40	
Error	23.29	72	0.32		

APPENDIX 4 (continued)

A. (continued)

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	P
Response (R)	0.02	1	0.02	0.01	
R x O	1.81	1	1.81	0.75	
Error	43.26	18	2.40		
P x R	0.73	1	0.73	1.46	
P x R x O	1.20	1	1.20	2.41	
Error	8.96	18	0.50		
Pr x R	2.75	1	2.75	3.60	
Pr x R x O	0.46	1	0.46	0.61	
Error	13.77	18	0.77		
P x Pr x R	1.11	1	1.11	3.51	
P x Pr x R x O	0.64	1	0.64	2.03	
Error	5.69	18	0.32		
H x R	0.01	1	0.01	0.03	
H x R x O	0.98	1	0.98	3.01	
Error	5.86	18	0.33		
P x H x R	0.66	1	0.66	1.68	
P x H x R x O	1.49	1	1.49	3.80	
Error	7.07	18	0.39		
Pr x H x R	6.18	1	6.18	33.80	< .001
Pr x H x R x O	0.42	1	0.42	2.30	
Error	3.29	18	0.18		
P x Pr x H x R	0.10	1	0.10	0.24	
P x Pr x H x R x O	0.05	1	0.05	0.14	
Error	7.26	18	0.40		
D x R	2.02	4	0.51	1.10	
D x R x O	2.42	4	0.61	1.32	
Error	32.97	72	0.46		
P x D x R	0.39	4	0.10	0.42	
P x D x R x O	2.21	4	0.55	2.37	
Error	16.81	72	0.23		
Pr x D x R	0.18	4	0.04	0.11	
Pr x D x R x O	2.38	4	0.60	1.55	
Error	27.62	72	0.38		
P x Pr x D x R	1.47	4	0.37	1.21	
P x Pr x D x R x O	0.77	4	0.19	0.64	
Error	21.76	72	0.30		
H x D x R	1.29	4	0.32	1.37	
H x D x R x O	1.74	4	0.44	1.84	
Error	17.03	72	0.24		
P x H x D x R	5.14	4	1.29	3.80	< .05
P x H x D x R x O	1.71	4	0.43	1.27	
Error	24.36	72	0.34		
Pr x H x D x R	3.52	4	0.88	3.00	< .05
Pr x H x D x R x O	0.95	4	0.24	0.81	
Error	21.11	72	0.29		
P x Pr x H x D x R	1.39	4	0.35	1.01	
P x Pr x H x D x R x O	0.19	4	0.05	0.14	
Error	24.69	72	0.34		

APPENDIX 4 (continued)

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 4.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	103406	1	103406	0.09	
Error	20487451	18	1138192		
Period (P)	2922	1	2922	0.06	
P x O	924	1	924	0.02	
Error	886663	18	49259		
Probability (Pr)	66462	1	66462	0.54	
Pr x O	172942	1	172942	1.40	
Error	2228664	18	123815		
P x Pr	1538	1	1538	0.04	
P x Pr x O	88	1	88	0.00	
Error	681965	18	37887		
Half (H)	164039	1	164039	6.68	< .05
H x O	11520	1	11520	0.47	
Error	441781	18	24543		
P x H	11326	1	11326	0.36	
P x H x O	1015	1	1015	0.03	
Error	566622	18	31479		
Pr x H	151105	1	151105	2.96	
Pr x H x O	93393	1	93393	1.83	
Error	918929	18	51052		
P x Pr x H	17117	1	17117	0.51	
P x Pr x H x O	25424	1	25424	0.76	
Error	601036	18	33391		
Discriminability (D)	7962678	4	1990670	28.72	< .001
D x O	248061	4	62015	0.89	
Error	4991376	72	69325		
P x D	35488	4	8872	0.19	
P x D x O	100643	4	25161	0.55	
Error	3321621	72	46134		
Pr x D	65584	4	16396	0.28	
Pr x D x O	209809	4	52453	0.91	
Error	4162379	72	57811		
P x Pr x D	261319	4	65330	1.20	
P x Pr x D x O	180069	4	45017	0.83	
Error	3904304	72	54227		
H x D	162379	4	40595	1.15	
H x D x O	115286	4	28822	0.82	
Error	2532444	72	35173		
P x H x D	52771	4	13193	0.35	
P x H x D x O	376139	4	94035	2.48	
Error	2727238	72	37878		
Pr x H x D	249197	4	62299	1.82	
Pr x H x D x O	175189	4	43797	1.28	
Error	2462293	72	34199		
P x Pr x H x D	433312	4	108328	3.63	< .01
P x Pr x H x D x O	73499	4	18375	0.62	
Error	2149812	72	29859		

APPENDIX 4 (continued)

B. (continued)

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Response (R)	135018	1	135018	0.75	
R x O	11224	1	11224	0.06	
Error	3236565	18	179809		
P x R	80564	1	80564	1.48	
P x R x O	4656	1	4656	0.09	
Error	982504	18	54584		
Pr x R	4365039	1	4365039	46.94	< .001
Pr x R x O	89777	1	89777	0.97	
Error	1673802	18	92989		
P x Pr x R	189480	1	189480	3.54	
P x Pr x R x O	3905	1	3905	0.07	
Error	963859	18	53548		
H x R	994	1	994	0.02	
H x R x O	143181	1	143181	2.55	
Error	1012071	18	56226		
P x H x R	63667	1	63667	1.55	
P x H x R x O	117407	1	117407	2.86	
Error	739701	18	41095		
Pr x H x R	645030	1	645030	31.78	< .001
Pr x H x R x O	45951	1	45951	2.26	
Error	365294	18	20294		
P x Pr x H x R	12822	1	12822	0.30	
P x Pr x H x R x O	7081	1	7081	0.16	
Error	776871	18	43160		
D x R	40885	4	10221	0.13	
D x R x O	63788	4	15947	0.21	
Error	5478265	72	76087		
P x D x R	58061	4	14515	0.36	
P x D x R x O	232619	4	58155	1.46	
Error	2872692	72	39899		
Pr x D x R	386661	4	96665	1.52	
Pr x D x R x O	325019	4	81255	1.28	
Error	4572344	72	63505		
P x Pr x D x R	89528	4	22382	0.70	
P x Pr x D x R x O	267435	4	66859	2.10	
Error	2289100	72	31793		
H x D x R	79272	4	19818	0.50	
H x D x R x O	192988	4	48247	1.23	
Error	2828519	72	39285		
P x H x D x R	563925	4	140981	3.67	< .01
P x H x D x R x O	40028	4	10007	0.26	
Error	2766635	72	38425		
Pr x H x D x R	182537	4	45634	1.50	
Pr x H x D x R x O	384107	4	96027	3.16	< .05
Error	2190917	72	30429		
P x Pr x H x D x R	37082	4	9270	0.28	
P x Pr x H x D x R x O	258375	4	64594	1.96	
Error	2372765	72	32955		

APPENDIX 4 (continued)

C. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 4.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	73.40	1	73.40	1.41	
Error	935.79	18	51.99		
Period (P)	0.18	1	0.18	0.11	
P x O	0.86	1	0.86	0.53	
Error	29.30	18	1.63		
Probability (Pr)	1.63	1	1.63	1.70	
Pr x O	1.43	1	1.43	1.50	
Error	17.25	18	0.96		
P x Pr	0.10	1	0.10	0.23	
P x Pr x O	0.35	1	0.35	0.83	
Error	7.70	18	0.43		
Half (H)	5.89	1	5.89	14.79	< .01
H x O	0.06	1	0.06	0.16	
Error	7.17	18	0.40		
P x H	0.77	1	0.77	1.41	
P x H x O	0.34	1	0.34	0.61	
Error	9.91	18	0.55		
Pr x H	2.04	1	2.04	5.07	< .05
Pr x H x O	2.42	1	2.42	6.00	< .05
Error	7.26	18	0.40		
P x Pr x H	0.31	1	0.31	1.27	
P x Pr x H x O	0.00	1	0.00	0.00	
Error	4.36	18	0.24		
Discriminability (D)	200.08	4	50.02	50.56	< .001
D x O	4.47	4	1.12	1.13	
Error	71.24	72	0.99		
P x D	3.51	4	0.88	2.82	< .05
P x D x O	0.51	4	0.13	0.41	
Error	22.39	72	0.31		
Pr x D	0.96	4	0.24	0.74	
Pr x D x O	0.92	4	0.23	0.71	
Error	23.19	72	0.32		
P x Pr x D	1.30	4	0.33	1.04	
P x Pr x D x O	1.46	4	0.36	1.17	
Error	22.48	72	0.31		
H x D	0.08	4	0.02	0.05	
H x D x O	0.29	4	0.07	0.19	
Error	27.92	72	0.39		
P x H x D	0.77	4	0.19	0.92	
P x H x D x O	1.17	4	0.29	1.39	
Error	15.10	72	0.21		
Pr x H x D	2.07	4	0.52	1.99	
Pr x H x D x O	0.20	4	0.05	0.19	
Error	18.74	72	0.26		
P x Pr x H x D	4.12	4	1.03	3.95	< .01
P x Pr x H x D x O	0.15	4	0.04	0.15	
Error	18.78	72	0.26		

APPENDIX 4 (continued)

C. (continued)

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Response (R)	2.11	1	2.11	0.37	
R x O	0.39	1	0.39	0.07	
Error	102.68	18	5.71		
P x R	1.93	1	1.93	2.87	
P x R x O	0.77	1	0.77	1.15	
Error	12.10	18	0.67		
Pr x R	13.01	1	13.01	9.09	< .01
Pr x R x O	0.12	1	0.12	0.08	
Error	25.76	18	1.43		
P x Pr x R	1.69	1	1.69	4.52	< .05
P x Pr x R x O	0.49	1	0.49	1.31	
Error	6.73	18	0.37		
H x R	0.06	1	0.06	0.24	
H x R x O	1.25	1	1.25	4.70	< .05
Error	4.80	18	0.27		
P x H x R	0.01	1	0.01	0.02	
P x H x R x O	0.00	1	0.00	0.00	
Error	7.06	18	0.39		
Pr x H x R	5.51	1	5.51	13.01	< .01
Pr x H x R x O	0.10	1	0.10	0.24	
Error	7.62	18	0.42		
P x Pr x H x R	0.01	1	0.01	0.04	
P x Pr x H x R x O	0.04	1	0.04	0.13	
Error	5.18	18	0.29		
D x R	0.45	4	0.11	0.26	
D x R x O	0.94	4	0.23	0.55	
Error	30.52	72	0.42		
P x D x R	1.48	4	0.37	1.13	
P x D x R x O	3.29	4	0.82	2.52	< .05
Error	23.50	72	0.33		
Pr x D x R	0.20	4	0.05	0.12	
Pr x D x R x O	2.65	4	0.66	1.56	
Error	30.66	72	0.43		
P x Pr x D x R	1.25	4	0.31	0.91	
P x Pr x D x R x O	0.61	4	0.15	0.44	
Error	24.69	72	0.34		
H x D x R	2.06	4	0.51	1.82	
H x D x R x O	0.64	4	0.16	0.57	
Error	20.42	72	0.28		
P x H x D x R	2.36	4	0.59	2.40	
P x H x D x R x O	0.28	4	0.07	0.29	
Error	17.65	72	0.25		
Pr x H x D x R	0.14	4	0.04	0.12	
Pr x H x D x R x O	2.09	4	0.52	1.77	
Error	21.28	72	0.30		
P x Pr x H x D x R	0.88	4	0.22	1.12	
P x Pr x H x D x R x O	2.25	4	0.56	2.86	< .05
Error	14.16	72	0.20		

APPENDIX 5 : EXPERIMENT 5. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 5.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Period (P)	0.09	1	0.09	0.84	
Error	3.10	29	0.11		
Probability (Pr)	0.23	1	0.23	1.92	
Error	3.45	29	0.12		
P x Pr	0.12	1	0.12	1.26	
Error	2.71	29	0.09		
Half (H)	0.11	1	0.11	0.97	
Error	3.28	29	0.11		
P x H	0.04	1	0.04	0.63	
Error	2.02	29	0.07		
Pr x H	0.01	1	0.01	0.16	
Error	2.65	29	0.09		
P x Pr x H	0.02	1	0.02	0.25	
Error	2.51	29	0.09		
Response (R)	0.00	1	0.00	0.00	
Error	20.06	29	0.69		
P x R	0.03	1	0.03	0.30	
Error	3.24	29	0.11		
Pr x R	0.12	1	0.12	0.51	
Error	6.77	29	0.23		
P x Pr x R	0.26	1	0.26	1.86	
Error	4.01	29	0.14		
H x R	0.37	1	0.37	3.30	
Error	3.25	29	0.11		
P x H x R	0.05	1	0.05	0.65	
Error	2.05	29	0.07		
Pr x H x R	5.16	1	5.16	41.18	< .001
Error	3.63	29	0.13		
P x Pr x H x R	0.02	1	0.02		
Error	3.17	29	0.11		

APPENDIX 5 (continued)

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 5.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Period (P)	50551	1	50551	2.69	
Error	545569	29	18813		
Probability (Pr)	359	1	359	0.03	
Error	320695	29	11058		
P x Pr	52553	1	52553	8.41	< .01
Error	181175	29	6247		
Half (H)	3611	1	3611	0.24	
Error	444242	29	15319		
P x H	383	1	383	0.04	
Error	304770	29	10509		
Pr x H	1986	1	1986	0.13	
Error	441993	29	15241		
P x Pr x H	26494	1	26494	2.01	
Error	381605	29	13159		
Response (R)	211751	1	211751	9.58	< .01
Error	640842	29	22099		
P x R	14833	1	14833	1.89	
Error	227001	29	7828		
Pr x R	89902	1	89902	6.19	< .05
Error	420870	29	14513		
P x Pr x R	9223	1	9223	1.50	
Error	178624	29	6159		
H x R	31100	1	31100	3.55	
Error	254364	29	8771		
P x H x R	12011	1	12011	1.48	
Error	234677	29	8092		
Pr x H x R	286451	1	286451	19.14	< .001
Error	434017	29	14966		
P x Pr x H x R	3357	1	3357	0.45	
Error	216509	29	7466		

APPENDIX 5 (continued)

C. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 5.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Period (P)	0.08	1	0.08	0.43	
Error	5.58	29	0.19		
Probability (Pr)	0.02	1	0.02	0.12	
Error	4.22	29	0.15		
P x Pr	0.01	1	0.01	0.09	
Error	3.60	29	0.12		
Half (H)	0.26	1	0.26	3.00	
Error	2.48	29	0.09		
P x H	0.21	1	0.21	2.13	
Error	2.89	29	0.10		
Pr x H	0.09	1	0.09	0.55	
Error	4.64	29	0.16		
P x Pr x H	0.01	1	0.01	0.02	
Error	4.00	29	0.14		
Response (R)	5.27	1	5.27	8.47	< .01
Error	18.05	29	0.62		
P x R	0.20	1	0.20	1.44	
Error	4.11	29	0.14		
Pr x R	0.78	1	0.78	4.60	< .05
Error	4.89	29	0.17		
P x Pr x R	0.03	1	0.03	0.40	
Error	1.84	29	0.06		
H x R	0.00	1	0.00	0.00	
Error	3.28	29	0.11		
P x H x R	0.02	1	0.02	0.20	
Error	2.57	29	0.09		
Pr x H x R	4.20	1	4.20	31.62	< .001
Error	3.85	29	0.13		
P x Pr x H x R	0.10	1	0.10	1.13	
Error	2.56	29	0.09		

APPENDIX 6 : EXPERIMENT 6. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 6.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	P
Order (O)	0.03	1	0.03	0.22	
Error	2.13	18	0.12		
Period (P)	0.12	1	0.12	1.32	
P x O	0.43	1	0.43	4.73	< .05
Error	1.64	18	0.09		
Context (C)	0.03	1	0.03	0.40	
C x O	0.13	1	0.13	1.68	
Error	1.38	18	0.08		
P x C	0.12	1	0.12	1.76	
P x C x O	0.09	1	0.09	1.28	
Error	1.19	18	0.07		
Half (H)	0.14	1	0.14	1.94	
H x O	0.07	1	0.07	0.92	
Error	1.31	18	0.07		
P x H	0.06	1	0.06	1.56	
P x H x O	0.01	1	0.01	0.39	
Error	0.64	18	0.04		
C x H	0.03	1	0.03	0.52	
C x H x O	0.03	1	0.03	0.52	
Error	1.03	18	0.06		
P x C x H	0.03	1	0.03	0.30	
P x C x H x O	0.07	1	0.07	0.74	
Error	1.79	18	0.10		
Response (R)	0.15	1	0.15	3.17	
R x O	0.01	1	0.01	0.14	
Error	0.85	18	0.05		
P x R	0.00	1	0.00	0.07	
P x R x O	0.00	1	0.00	0.07	
Error	0.93	18	0.05		
C x R	0.16	1	0.16	1.86	
C x R x O	0.00	1	0.00	0.03	
Error	1.53	18	0.08		
P x C x R	0.14	1	0.14	3.16	
P x C x R x O	0.01	1	0.01	0.11	
Error	0.82	18	0.05		
H x R	0.08	1	0.08	2.59	
H x R x O	0.45	1	0.45	14.13	< .01
Error	0.57	18	0.03		
P x H x R	0.01	1	0.01	0.02	
P x H x R x O	0.00	1	0.00	0.00	
Error	1.23	18	0.07		
C x H x R	0.03	1	0.03	0.56	
C x H x R x O	0.12	1	0.12	2.04	
Error	1.09	18	0.06		
P x C x H x R	0.00	1	0.00	0.00	
P x C x H x R x O	0.16	1	0.16	1.83	
Error	1.58	18	0.09		

APPENDIX 6 (continued)

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 6.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	3460	1	3460	0.03	
Error	2417860	18	134326		
Period (P)	24934	1	24934	2.12	
P x O	1722	1	1722	0.15	
Error	211820	18	11768		
Context (C)	1754	1	1754	0.12	
C x O	5113	1	5113	0.35	
Error	266179	18	14788		
P x C	1802	1	1802	0.18	
P x C x O	3442	1	3442	0.34	
Error	180767	18	10043		
Half (H)	47	1	47	0.01	
H x O	20046	1	20046	3.48	
Error	103791	18	5766		
P x H	3396	1	3396	0.66	
P x H x O	6020	1	6020	1.17	
Error	93003	18	5167		
C x H	4	1	4	0.00	
C x H x O	743	1	743	0.29	
Error	45545	18	2530		
P x C x H	7380	1	7380	1.38	
P x C x H x O	310	1	310	0.06	
Error	96100	18	5339		
Response (R)	141074	1	141074	10.37	< .01
R x O	14907	1	14907	1.10	
Error	244783	18	13599		
P x R	1175	1	1175	0.48	
P x R x O	1446	1	1446	0.59	
Error	43870	18	2437		
C x R	873	1	873	0.21	
C x R x O	2049	1	2049	0.50	
Error	73597	18	4089		
P x C x R	3726	1	3726	0.63	
P x C x R x O	3025	1	3025	0.51	
Error	106715	18	5929		
H x R	4923	1	4923	2.81	
H x R x O	4254	1	4254	2.43	
Error	31500	18	1750		
P x H x R	493	1	493	0.12	
P x H x R x O	519	1	519	0.12	
Error	76683	18	4260		
C x H x R	650	1	650	0.18	
C x H x R x O	101	1	101	0.03	
Error	65350	18	3631		
P x C x H x R	117	1	117	0.03	
P x C x H x R x O	9856	1	9856	2.65	
Error	66950	18	3719		

APPENDIX 6 (continued)

C. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 6.

SOURCE	SUM OF SQUARES	df	MEAN SQUARES	F	p
Order (O)	0.16	1	0.16	0.05	
Error	62.35	18	3.46		
Period (P)	0.15	1	0.15	0.23	
P x O	0.01	1	0.01	0.02	
Error	11.65	18	0.65		
Context (C)	0.01	1	0.01	0.01	
C x O	0.01	1	0.01	0.01	
Error	13.93	18	0.77		
P x C	0.25	1	0.25	0.44	
P x C x O	0.00	1	0.00	0.00	
Error	10.19	18	0.57		
Half (H)	0.01	1	0.01	0.10	
H x O	0.00	1	0.00	0.04	
Error	1.65	18	0.09		
P x H	0.00	1	0.00	0.00	
P x H x O	0.09	1	0.09	0.82	
Error	2.09	18	0.12		
C x H	0.00	1	0.00	0.03	
C x H x O	0.08	1	0.08	0.92	
Error	1.59	18	0.09		
P x C x H	0.01	1	0.01	0.15	
P x C x H x O	0.01	1	0.01	0.07	
Error	1.32	18	0.07		
Response (R)	2.03	1	2.03	3.18	
R x O	0.36	1	0.36	0.56	
Error	11.52	18	0.64		
P x R	0.06	1	0.06	0.93	
P x R x O	0.01	1	0.01	0.06	
Error	1.12	18	0.06		
C x R	0.44	1	0.44	2.48	
C x R x O	0.01	1	0.01	0.06	
Error	3.22	18	0.18		
P x C x R	0.02	1	0.02	0.10	
P x C x R x O	0.13	1	0.13	0.71	
Error	3.35	18	0.19		
H x R	0.23	1	0.23	6.07	< .05
H x R x O	0.01	1	0.01	0.19	
Error	0.67	18	0.04		
P x H x R	0.13	1	0.13	4.98	< .05
P x H x R x O	0.01	1	0.01	0.10	
Error	0.48	18	0.03		
C x H x R	0.01	1	0.01	0.01	
C x H x R x O	0.02	1	0.02	0.18	
Error	1.63	18	0.09		
P x C x H x R	0.08	1	0.08	1.60	
P x C x H x R x O	0.32	1	0.32	6.28	< .05
Error	0.91	18	0.05		

APPENDIX 7 : EXPERIMENT 7. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 7.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Period (P)	0.11	1	0.11	3.09	
Error	0.33	9	0.04		
Context (C)	0.01	1	0.01	0.30	
Error	0.40	9	0.04		
P x C	0.04	1	0.04	0.51	
Error	0.78	9	0.09		
Half (H)	0.00	1	0.00	0.00	
Error	0.51	9	0.06		
P x H	0.38	1	0.38	8.64	< .05
Error	0.39	9	0.04		
C x H	0.07	1	0.07	4.74	
Error	0.13	9	0.02		
P x C x H	0.00	1	0.00	0.00	
Error	0.35	9	0.04		
Response (R)	0.59	1	0.59	3.82	
Error	1.40	9	0.16		
P x R	0.02	1	0.02	0.22	
Error	0.67	9	0.07		
C x R	0.04	1	0.04	0.23	
Error	1.53	9	0.17		
P x C x R	0.04	1	0.04	0.39	
Error	0.91	9	0.10		
H x R	0.32	1	0.32	4.19	
Error	0.68	9	0.08		
P x H x R	0.01	1	0.01	0.01	
Error	0.80	9	0.09		
C x H x R	0.06	1	0.06	1.31	
Error	0.39	9	0.04		
P x C x H x R	0.01	1	0.01	0.13	
Error	0.21	9	0.02		

APPENDIX 7 (continued)

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 7.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Period (P)	6737	1	6737	0.37	
Error	162590	9	18066		
Context (C)	11350	1	11350	0.52	
Error	196432	9	21826		
P x C	658	1	658	0.08	
Error	73396	9	8155		
Half (H)	149	1	149	0.03	
Error	49364	9	5485		
P x H	4644	1	4644	0.80	
Error	69335	9	7034		
C x H	90	1	90	0.01	
Error	97653	9	10850		
P x C x H	2297	1	2297	0.32	
Error	65608	9	7290		
Response (R)	68525	1	68525	2.26	
Error	273171	9	30352		
P x R	4955	1	4955	1.48	
Error	30161	9	3351		
C x R	47838	1	47838	4.18	
Error	102910	9	11434		
P x C x R	219	1	219	0.07	
Error	30014	9	3335		
H x R	52	1	52	0.01	
Error	56964	9	6329		
P x H x R	18632	1	18632	2.13	
Error	78836	9	8759		
C x H x R	43573	1	43573	12.29	< .01
Error	31905	9	3545		
P x C x H x R	2208	1	2208	0.30	
Error	66741	9	7416		

APPENDIX 7 (continued)

C. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 7.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Period (P)	0.03	1	0.03	0.13	
Error	2.29	9	0.25		
Context (C)	0.18	1	0.18	1.38	
Error	1.15	9	0.13		
P x C	0.18	1	0.18	0.89	
Error	1.77	9	0.20		
Half (H)	0.06	1	0.06	1.15	
Error	0.47	9	0.05		
P x H	0.01	1	0.01	0.06	
Error	1.56	9	0.17		
C x H	0.01	1	0.01	0.04	
Error	0.70	9	0.08		
P x C x H	0.01	1	0.01	0.12	
Error	0.39	9	0.04		
Response (R)	3.45	1	3.45	4.42	
Error	7.03	9	0.78		
P x R	0.02	1	0.02	0.34	
Error	0.59	9	0.07		
C x R	0.33	1	0.33	1.69	
Error	1.78	9	0.20		
P x C x R	0.13	1	0.13	1.28	
Error	0.89	9	0.10		
H x R	0.03	1	0.03	0.95	
Error	0.31	9	0.04		
P x H x R	0.03	1	0.03	0.73	
Error	0.41	9	0.05		
C x H x R	0.16	1	0.16	10.89	< .01
Error	0.13	9	0.02		
P x C x H x R	0.00	1	0.00	0.00	
Error	0.32	9	0.04		

APPENDIX 8 : EXPERIMENT 8. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 8.

SOURCE	SUM OF SQUARES	df	MEAN SQUARES	F	p
Order (O)	0.05	1	0.05	0.03	
Error	33.79	18	1.88		
Period (P)	0.02	1	0.02	0.22	
P x O	0.01	1	0.01	0.09	
Error	1.75	18	0.10		
Set (S)	0.07	1	0.07	0.69	
S x O	0.68	1	0.68	7.12	< .05
Error	1.71	18	0.10		
P x S	0.02	1	0.02	0.11	
P x S x O	0.50	1	0.50	2.77	
Error	3.27	18	0.18		
Half (H)	0.01	1	0.01	0.08	
H x O	0.00	1	0.00	0.00	
Error	2.10	18	0.12		
P x H	0.19	1	0.19	2.51	
P x H x O	0.02	1	0.02	0.27	
Error	1.38	18	0.08		
S x H	0.04	1	0.04	0.22	
S x H x O	0.01	1	0.01	0.07	
Error	2.92	18	0.16		
P x S x H	0.01	1	0.01	0.02	
P x S x H x O	0.01	1	0.01	0.09	
Error	2.20	18	0.12		
Discriminability (D)	94.24	4	23.56	65.30	< .001
D x O	2.02	4	0.50	1.40	
Error	25.98	72	0.36		
P x D	0.09	4	0.02	0.20	
P x D x O	0.22	4	0.05	0.48	
Error	8.26	72	0.11		
S x D	0.38	4	0.10	1.06	
S x D x O	0.33	4	0.08	0.93	
Error	6.51	72	0.09		
P x S x D	0.15	4	0.04	0.34	
P x S x D x O	1.48	4	0.37	3.27	< .05
Error	8.15	72	0.11		
H x D	0.26	4	0.06	0.69	
H x D x O	0.26	4	0.06	0.68	
Error	6.78	72	0.09		
P x H x D	1.15	4	0.29	3.18	< .05
P x H x D x O	0.04	4	0.01	0.11	
Error	6.51	72	0.09		
S x H x D	0.12	4	0.03	0.37	
S x H x D x O	0.13	4	0.03	0.41	
Error	5.80	72	0.08		
P x S x H x D	0.39	4	0.10	1.29	
P x S x H x D x O	0.14	4	0.04	0.48	
Error	5.38	72	0.07		

APPENDIX 8 (continued)

A. (continued)

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Response (R)	0.65	1	0.65	1.47	
R x O	0.01	1	0.01	0.01	
Error	8.01	18	0.45		
P x R	0.00	1	0.00	0.00	
P x R x O	0.15	1	0.15	0.65	
Error	4.20	18	0.23		
S x R	4.59	1	4.59	2.39	
S x R x O	0.99	1	0.99	0.51	
Error	34.52	18	1.92		
P x S x R	0.03	1	0.03	0.25	
P x S x R x O	0.15	1	0.15	1.11	
Error	2.50	18	0.14		
H x R	0.39	1	0.39	2.24	
H x R x O	0.15	1	0.15	0.83	
Error	3.16	18	0.18		
P x H x R	0.31	1	0.31	3.25	
P x H x R x O	0.01	1	0.01	0.07	
Error	1.70	18	0.09		
S x H x R	0.00	1	0.00	0.00	
S x H x R x O	0.01	1	0.01	0.19	
Error	1.06	18	0.06		
P x S x H x R	0.08	1	0.08	0.90	
P x S x H x R x O	0.35	1	0.35	3.78	
Error	1.68	18	0.09		
D x R	3.67	4	0.92	3.62	< .01
D x R x O	0.08	4	0.02	0.08	
Error	18.29	72	0.25		
P x D x R	0.17	4	0.04	0.32	
P x D x R x O	0.27	4	0.07	0.52	
Error	9.39	72	0.13		
S x D x R	1.46	4	0.37	1.48	
S x D x R x O	0.25	4	0.06	0.25	
Error	17.73	72	0.25		
P x S x D x R	0.44	4	0.11	1.15	
P x S x D x R x O	0.71	4	0.18	1.85	
Error	6.90	72	0.10		
H x D x R	1.66	4	0.42	4.59	< .01
H x D x R x O	2.19	4	0.55	6.04	< .001
Error	6.51	72	0.09		
P x H x D x R	1.44	4	0.36	3.91	< .01
P x H x D x R x O	0.20	4	0.05	0.54	
Error	6.62	72	0.09		
S x H x D x R	0.31	4	0.08	0.89	
S x H x D x R x O	0.19	4	0.05	0.56	
Error	6.22	72	0.09		
P x S x H x D x R	0.78	4	0.20	2.15	
P x S x H x D x R x O	0.69	4	0.17	1.89	
Error	6.56	72	0.09		

APPENDIX 8 (continued)

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 8.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	977	1	977	0.00	
Error	63383498	18	3521305		
Period (P)	103673	1	103673	3.83	
P x O	4387	1	4387	0.16	
Error	486611	18	27033		
Set (S)	33544	1	33544	0.86	
S x O	6831	1	6831	0.18	
Error	698947	18	38830		
P x S	10250	1	10250	1.09	
P x S x O	39741	1	39741	4.21	
Error	169879	18	9438		
Half (H)	1270	1	1270	0.05	
H x O	34045	1	34045	1.42	
Error	431506	18	23973		
P x H	5842	1	5842	0.23	
P x H x O	3757	1	3757	0.15	
Error	465562	18	25864		
S x H	22323	1	22323	0.45	
S x H x O	4390	1	4390	0.09	
Error	883110	18	49062		
P x S x H	3569	1	3569	0.18	
P x S x H x O	1535	1	1535	0.08	
Error	361272	18	20071		
Discriminability (D)	15724935	4	3931234	40.83	< .001
D x O	40490	4	10122	0.11	
Error	6933052	72	96292		
P x D	87186	4	21796	2.05	
P x D x O	15418	4	3855	0.36	
Error	766898	72	10651		
S x D	141289	4	35322	2.86	< .05
S x D x O	199301	4	49825	4.04	< .01
Error	887844	72	12331		
P x S x D	40345	4	10086	0.61	
P x S x D x O	111668	4	27917	1.69	
Error	1190137	72	16530		
H x D	119533	4	29883	1.85	
H x D x O	62246	4	15561	0.97	
Error	1160831	72	16123		
P x H x D	92262	4	23065	1.27	
P x H x D x O	77380	4	19345	1.07	
Error	1305404	72	18131		
S x H x D	50344	4	12586	0.71	
S x H x D x O	329278	4	82319	4.66	< .01
Error	1271340	72	17658		
P x S x H x D	54798	4	13700	0.69	
P x S x H x D x O	43538	4	10885	0.55	
Error	1424311	72	19782		

APPENDIX 8 (continued)

B. (continued)

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	P
Response (R)	371963	1	371963	3.67	
R x O	8810	1	8810	0.09	
Error	1823896	18	101328		
P x R	445	1	445	0.03	
P x R x O	14684	1	14684	0.89	
Error	298069	18	16559		
S x R	5022451	1	5022451	10.05	< .01
S x R x O	514824	1	514824	1.03	
Error	8998210	18	499901		
P x S x R	9895	1	9895	0.46	
P x S x R x O	14139	1	14139	0.66	
Error	383605	18	21311		
H x R	65624	1	65624	2.57	
H x R x O	17278	1	17278	0.68	
Error	460077	18	25560		
P x H x R	671	1	671	0.04	
P x H x R x O	20671	1	20671	1.20	
Error	310403	18	17245		
S x H x R	8222	1	8222	0.57	
S x H x R x O	83	1	83	0.01	
Error	260797	18	14489		
P x S x H x R	414	1	414	0.02	
P x S x H x R x O	66829	1	66829	3.25	
Error	370699	18	20594		
D x R	36843	4	9211	0.30	
D x R x O	34983	4	8746	0.29	
Error	2203723	72	30607		
P x D x R	117060	4	29265	2.24	
P x D x R x O	180818	4	45204	3.46	< .05
Error	940342	72	13060		
S x D x R	90635	4	22659	1.13	
S x D x R x O	43724	4	10931	0.54	
Error	1447718	72	20107		
P x S x D x R	91631	4	22908	1.87	
P x S x D x R x O	42200	4	10550	0.86	
Error	883831	72	12275		
H x D x R	70719	4	17680	1.21	
H x D x R x O	217998	4	54499	3.74	< .01
Error	1047823	72	14553		
P x H x D x R	8686	4	2172	0.12	
P x H x D x R x O	27172	4	6793	0.38	
Error	1273584	72	17689		
S x H x D x R	27701	4	6925	0.40	
S x H x D x R x O	24180	4	6045	0.35	
Error	1248126	72	17335		
P x S x H x D x R	69615	4	17404	0.86	
P x S x H x D x R x O	26038	4	6510	0.32	
Error	1457129	72	20238		

APPENDIX 8 (continued)

C. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 8.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	0.31	1	0.31	0.01	
Error	854.28	18	47.46		
Period (P)	2.31	1	2.31	5.15	< .05
P x O	0.34	1	0.34	0.76	
Error	8.06	18	0.45		
Set (S)	0.01	1	0.01	0.01	
S x O	0.09	1	0.09	0.29	
Error	5.63	18	0.31		
P x S	0.01	1	0.01	0.01	
P x S x O	0.41	1	0.41	1.45	
Error	5.06	18	0.28		
Half (H)	0.98	1	0.98	2.18	
H x O	0.00	1	0.00	0.00	
Error	8.12	18	0.45		
P x H	0.10	1	0.10	0.52	
P x H x O	0.11	1	0.11	0.57	
Error	3.39	18	0.19		
S x H	0.39	1	0.39	0.92	
S x H x O	0.23	1	0.23	0.53	
Error	7.66	18	0.43		
P x S x H	0.10	1	0.10	0.33	
P x S x H x O	0.01	1	0.01	0.04	
Error	5.28	18	0.29		
Discriminability (D)	134.66	4	33.66	36.73	< .001
D x O	7.51	4	1.88	2.05	
Error	65.98	72	0.92		
P x D	1.75	4	0.44	2.81	< .01
P x D x O	0.60	4	0.15	0.96	
Error	11.18	72	0.16		
S x D	0.32	4	0.08	0.32	
S x D x O	0.87	4	0.22	0.88	
Error	17.77	72	0.25		
P x S x D	0.12	4	0.03	0.16	
P x S x D x O	0.74	4	0.18	0.96	
Error	13.89	72	0.19		
H x D	0.90	4	0.23	1.30	
H x D x O	0.54	4	0.13	0.78	
Error	12.45	72	0.17		
P x H x D	0.36	4	0.09	0.51	
P x H x D x O	0.33	4	0.08	0.46	
Error	12.64	72	0.18		
S x H x D	0.10	4	0.03	0.25	
S x H x D x O	0.50	4	0.12	1.17	
Error	7.64	72	0.11		
P x S x H x D	0.28	4	0.07	0.39	
P x S x H x D x O	0.62	4	0.15	0.86	
Error	12.89	72	0.18		

APPENDIX 8 (continued)

C. (continued)

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Response (R)	20.73	1	20.73	8.25	< .05
R x O	0.20	1	0.20	0.08	
Error	45.22	18	2.51		
P x R	0.01	1	0.01	0.00	
P x R x O	0.73	1	0.73	2.64	
Error	4.95	18	0.27		
S x R	0.62	1	0.62	0.14	
S x R x O	0.16	1	0.16	0.03	
Error	81.45	18	4.53		
P x S x R	0.01	1	0.01	0.01	
P x S x R x O	0.69	1	0.69	1.61	
Error	7.70	18	0.43		
H x R	0.04	1	0.04	0.17	
H x R x O	0.19	1	0.19	0.74	
Error	4.62	18	0.26		
P x H x R	0.62	1	0.62	2.55	
P x H x R x O	0.04	1	0.04	0.15	
Error	4.40	18	0.24		
S x H x R	0.75	1	0.75	3.18	
S x H x R x O	0.00	1	0.00	0.00	
Error	4.23	18	0.24		
P x S x H x R	0.01	1	0.01	0.02	
P x S x H x R x O	0.55	1	0.55	2.27	
Error	4.34	18	0.24		
D x R	1.72	4	0.43	2.43	
D x R x O	0.29	4	0.07	0.40	
Error	12.76	72	0.18		
P x D x R	0.39	4	0.10	0.56	
P x D x R x O	0.59	4	0.15	0.85	
Error	12.41	72	0.17		
S x D x R	0.63	4	0.16	0.74	
S x D x R x O	0.48	4	0.12	0.56	
Error	15.51	72	0.22		
P x S x D x R	0.16	4	0.04	0.32	
P x S x D x R x O	0.96	4	0.24	1.92	
Error	8.94	72	0.12		
H x D x R	0.93	4	0.23	1.46	
H x D x R x O	1.58	4	0.40	2.49	
Error	11.44	72	0.16		
P x H x D x R	0.86	4	0.21	2.11	
P x H x D x R x O	0.19	4	0.05	0.47	
Error	7.37	72	0.10		
S x H x D x R	0.48	4	0.12	0.82	
S x H x D x R x O	0.99	4	0.25	1.70	
Error	10.57	72	0.15		
P x S x H x D x R	0.10	4	0.03	0.19	
P x S x H x D x R x O	1.52	4	0.38	2.83	< .05
Error	9.71	72	0.13		

APPENDIX 9 : EXPERIMENT 9. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 9.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	P
Order (O)	0.03	1	0.03	0.11	
Error	5.41	18	0.30		
Period (P)	0.05	1	0.05	0.87	
P x O	0.01	1	0.01	0.16	
Error	1.11	18	0.06		
Set (S)	0.01	1	0.01	0.03	
S x O	0.00	1	0.00	0.00	
Error	2.69	18	0.15		
P x S	0.14	1	0.14	1.75	
P x S x O	0.04	1	0.04	0.46	
Error	1.45	18	0.08		
Half (H)	0.16	1	0.16	9.26	< .01
H x O	0.01	1	0.01	0.54	
Error	0.31	18	0.02		
P x H	0.01	1	0.01	0.30	
P x H x O	0.10	1	0.10	4.13	
Error	0.44	18	0.02		
S x H	0.02	1	0.02	0.15	
S x H x O	0.07	1	0.07	0.67	
Error	1.84	18	0.10		
P x S x H	0.01	1	0.01	0.28	
P x S x H x O	0.00	1	0.00	0.00	
Error	0.70	18	0.04		
Response (R)	0.01	1	0.01	0.06	
R x O	0.00	1	0.00	0.01	
Error	4.24	18	0.24		
P x R	0.18	1	0.18	3.78	
P x R x O	0.01	1	0.01	0.07	
Error	0.85	18	0.05		
S x R	3.43	1	3.43	17.11	< .001
S x R x O	0.16	1	0.16	0.78	
Error	3.61	18	0.20		
P x S x R	0.01	1	0.01	0.03	
P x S x R x O	0.01	1	0.01	0.11	
Error	1.33	18	0.07		
H x R	0.08	1	0.08	1.93	
H x R x O	0.01	1	0.01	0.30	
Error	0.71	18	0.04		
P x H x R	0.01	1	0.01	0.05	
P x H x R x O	0.08	1	0.08	1.56	
Error	0.92	18	0.05		
S x H x R	0.04	1	0.04	0.42	
S x H x R x O	0.07	1	0.07	0.78	
Error	1.55	18	0.09		
P x S x H x R	0.00	1	0.00	0.01	
P x S x H x R x O	0.16	1	0.16	2.28	
Error	1.27	18	0.07		

APPENDIX 9 (continued)

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 9.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	P
Order (O)	36191	1	36191	0.17	
Error	3896941	18	216497		
Period (P)	24467	1	24467	1.96	
P x O	5421	1	5421	0.43	
Error	224779	18	12488		
Set (S)	24310	1	24310	1.73	
S x O	22	1	22	0.00	
Error	252405	18	14022		
P x S	283	1	283	0.02	
P x S x O	16790	1	16790	1.16	
Error	261602	18	14533		
Half (H)	221	1	221	0.03	
H x O	4025	1	4025	0.49	
Error	146940	18	8163		
P x H	1425	1	1425	0.30	
P x H x O	15281	1	15281	3.23	
Error	85206	18	4734		
S x H	7969	1	7969	0.57	
S x H x O	37843	1	37843	2.69	
Error	252847	18	14047		
P x S x H	1550	1	1550	0.18	
P x S x H x O	43863	1	43863	5.18	< .05
Error	152350	18	8464		
Response (R)	179916	1	179916	3.51	
R x O	41996	1	41996	0.82	
Error	922066	18	51226		
P x R	41973	1	41973	6.83	< .05
P x R x O	196	1	196	0.03	
Error	110636	18	6146		
S x R	1682826	1	1682826	20.42	< .001
S x R x O	7710	1	7710	0.09	
Error	1483397	18	82410		
P x S x R	2977	1	2977	0.63	
P x S x R x O	38196	1	38196	8.07	< .05
Error	85227	18	4735		
H x R	843	1	843	0.10	
H x R x O	248	1	248	0.03	
Error	155934	18	8663		
P x H x R	24	1	24	0.01	
P x H x R x O	5711	1	5711	1.29	
Error	79556	18	4420		
S x H x R	973	1	973	0.09	
S x H x R x O	2586	1	2586	0.25	
Error	185552	18	10308		
P x S x H x R	1865	1	1865	0.42	
P x S x H x R x O	992	1	992	0.23	
Error	79200	18	4400		

APPENDIX 9 (continued)

C. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 9.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	13.61	1	13.61	1.90	
Error	128.99	18	7.17		
Period (P)	1.44	1	1.44	4.39	
P x O	0.75	1	0.75	2.30	
Error	5.90	18	0.33		
Set (S)	0.03	1	0.03	0.21	
S x O	0.73	1	0.73	4.97	< .05
Error	2.65	18	0.15		
P x S	0.02	1	0.02	0.32	
P x S x O	0.02	1	0.02	0.21	
Error	1.36	18	0.08		
Half (H)	0.00	1	0.00	0.00	
H x O	0.00	1	0.00	0.00	
Error	1.98	18	0.11		
P x H	0.01	1	0.01	0.05	
P x H x O	0.13	1	0.13	2.26	
Error	1.07	18	0.06		
S x H	0.08	1	0.08	1.28	
S x H x O	0.06	1	0.06	1.01	
Error	1.14	18	0.06		
P x S x H	0.03	1	0.03	0.45	
P x S x H x O	0.08	1	0.08	1.18	
Error	1.27	18	0.07		
Response (R)	1.09	1	1.09	0.79	
R x O	2.90	1	2.90	2.11	
Error	24.67	18	1.37		
P x R	0.65	1	0.65	6.46	< .05
P x R x O	0.01	1	0.01	0.03	
Error	1.80	18	0.10		
S x R	0.05	1	0.05	0.06	
S x R x O	0.06	1	0.06	0.08	
Error	13.40	18	0.75		
P x S x R	0.06	1	0.06	0.35	
P x S x R x O	0.00	1	0.00	0.00	
Error	3.25	18	0.18		
H x R	0.09	1	0.09	0.77	
H x R x O	0.01	1	0.01	0.12	
Error	2.07	18	0.12		
P x H x R	0.10	1	0.10	1.42	
P x H x R x O	0.75	1	0.75	10.48	< .01
Error	1.30	18	0.07		
S x H x R	0.02	1	0.02	0.32	
S x H x R x O	0.01	1	0.01	0.12	
Error	1.17	18	0.07		
P x S x H x R	0.46	1	0.46	5.72	< .05
P x S x H x R x O	0.01	1	0.01	0.06	
Error	1.46	18	0.08		

APPENDIX 10 : EXPERIMENT 10. ANALYSES OF VARIANCE:

A. Summary table of the analysis of variance of the percentage of stimuli for which responses were correct from Experiment 10.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	P
Order (O)	5.24	1	5.24	1.75	
Error	53.97	18	3.00		
Period (P)	0.12	1	0.12	0.32	
P x O	0.05	1	0.05	0.14	
Error	6.53	18	0.36		
Set (S)	36.31	1	36.31	25.20	< .001
S x O	1.51	1	1.51	1.04	
Error	25.94	18	1.44		
P x S	0.09	1	0.09	0.46	
P x S x O	0.77	1	0.77	3.98	
Error	3.48	18	0.19		
Half (H)	0.09	1	0.09	0.44	
H x O	0.01	1	0.01	0.01	
Error	3.51	18	0.20		
P x H	0.03	1	0.03	0.24	
P x H x O	0.01	1	0.01	0.00	
Error	2.52	18	0.14		
S x H	0.01	1	0.01	0.02	
S x H x O	1.99	1	1.99	6.75	< .05
Error	5.31	18	0.29		
P x S x H	0.73	1	0.73	2.14	
P x S x H x O	0.24	1	0.24	0.70	
Error	6.12	18	0.34		
Discriminability (D)	103.65	4	25.91	100.54	< .001
D x O	0.49	4	0.12	0.48	
Error	18.56	72	0.26		
P x D	0.91	4	0.23	1.25	
P x D x O	0.41	4	0.10	0.56	
Error	13.10	72	0.18		
S x D	2.88	4	0.72	3.21	< .05
S x D x O	1.39	4	0.35	1.54	
Error	16.17	72	0.22		
P x S x D	0.62	4	0.15	1.09	
P x S x D x O	1.63	4	0.41	2.88	< .05
Error	10.19	72	0.14		
H x D	0.07	4	0.02	0.10	
H x D x O	0.60	4	0.15	0.89	
Error	12.08	72	0.17		
P x H x D	0.35	4	0.09	0.41	
P x H x D x O	0.05	4	0.01	0.06	
Error	15.68	72	0.22		
S x H x D	0.21	4	0.05	0.29	
S x H x D x O	0.60	4	0.15	0.81	
Error	13.38	72	0.19		
P x S x H x D	0.98	4	0.25	1.47	
P x S x H x D x O	0.53	4	0.13	0.79	
Error	12.01	72	0.17		

APPENDIX 10 (continued)

A. (continued)

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Response (R)	1.06	1	1.06	0.96	
R x O	3.02	1	3.02	2.74	
Error	19.85	18	1.11		
P x R	0.22	1	0.22	2.38	
P x R x O	0.13	1	0.13	1.43	
Error	1.66	18	0.09		
S x R	0.52	1	0.52	1.19	
S x R x O	0.06	1	0.06	0.14	
Error	7.90	18	0.44		
P x S x R	0.01	1	0.01	0.02	
P x S x R x O	0.02	1	0.02	0.09	
Error	3.27	18	0.18		
H x R	0.24	1	0.24	0.82	
H x R x O	0.02	1	0.02	0.07	
Error	5.21	18	0.29		
P x H x R	0.01	1	0.01	0.04	
P x H x R x O	0.38	1	0.38	1.97	
Error	3.47	18	0.19		
S x H x R	0.15	1	0.15	0.79	
S x H x R x O	0.31	1	0.31	1.68	
Error	3.32	18	0.18		
P x S x H x R	0.13	1	0.13	1.20	
P x S x H x R x O	0.14	1	0.14	1.26	
Error	1.96	18	0.11		
D x R	2.06	4	0.52	1.78	
D x R x O	1.25	4	0.31	1.08	
Error	20.86	72	0.29		
P x D x R	0.96	4	0.24	1.31	
P x D x R x O	0.73	4	0.18	0.99	
Error	13.18	72	0.18		
S x D x R	1.89	4	0.47	2.85	< .05
S x D x R x O	1.22	4	0.31	1.85	
Error	11.92	72	0.17		
P x S x D x R	0.36	4	0.09	0.51	
P x S x D x R x O	0.41	4	0.10	0.59	
Error	12.45	72	0.17		
H x D x R	0.46	4	0.12	0.66	
H x D x R x O	0.59	4	0.15	0.85	
Error	12.60	72	0.18		
P x H x D x R	0.12	4	0.03	0.14	
P x H x D x R x O	0.98	4	0.24	1.16	
Error	15.12	72	0.21		
S x H x D x R	0.34	4	0.08	0.57	
S x H x D x R x O	0.35	4	0.09	0.59	
Error	10.74	72	0.15		
P x S x H x D x R	0.69	4	0.17	1.27	
P x S x H x D x R x O	1.81	4	0.45	3.35	< .05
Error	9.75	72	0.14		

APPENDIX 10 (continued)

B. Summary table of the analysis of variance of mean times for correct responses from Experiment 10.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	P
Order (O)	1455600	1	1455600	1.10	
Error	23780897	18	1321161		
Period (P)	93057	1	93057	0.96	
P x O	363329	1	363329	3.74	
Error	1747150	18	97064		
Set (S)	40533564	1	40533564	34.90	< .001
S x O	6386	1	6386	0.01	
Error	20905378	18	1161410		
P x S	78960	1	78960	1.68	
P x S x O	3924	1	3924	0.08	
Error	846906	18	47050		
Half (H)	32392	1	32392	0.71	
H x O	72597	1	72597	1.59	
Error	820097	18	45561		
P x H	6323	1	6323	0.11	
P x H x O	69162	1	69162	1.20	
Error	1036627	18	57590		
S x H	46754	1	46754	1.06	
S x H x O	45040	1	45040	1.02	
Error	797104	18	44284		
P x S x H	237761	1	237761	4.38	
P x S x H x O	7596	1	7596	0.14	
Error	976219	18	54234		
Discriminability (D)	11648003	4	2912001	54.87	< .001
D x O	199108	4	49777	0.94	
Error	3821391	72	53075		
P x D	62589	4	15647	0.94	
P x D x O	98866	4	24717	1.48	
Error	1203781	72	16719		
S x D	3451930	4	862983	20.88	< .001
S x D x O	20514	4	5129	0.12	
Error	2975631	72	41328		
P x S x D	179172	4	44793	2.21	
P x S x D x O	61206	4	15302	0.75	
Error	1461165	72	20294		
H x D	76069	4	19017	0.73	
H x D x O	109462	4	27366	1.05	
Error	1884122	72	26168		
P x H x D	85926	4	21481	1.56	
P x H x D x O	62561	4	15640	1.13	
Error	992483	72	13785		
S x H x D	13836	4	3459	0.15	
S x H x D x O	30587	4	7647	0.34	
Error	1616197	72	22447		
P x S x H x D	20362	4	5090	0.29	
P x S x H x D x O	234713	4	58678	3.29	< .05
Error	1282771	72	17816		

APPENDIX 10 (continued)

B. (continued)

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Response (R)	162	1	162	0.00	
R x O	21129	1	21129	0.29	
Error	1324860	18	73603		
P x R	6168	1	6168	0.21	
P x R x O	58895	1	58895	2.02	
Error	524274	18	29126		
S x R	61726	1	61726	1.54	
S x R x O	3869	1	3869	0.10	
Error	722842	18	40158		
P x S x R	15	1	15	0.00	
P x S x R x O	5604	1	5604	0.26	
Error	389587	18	21644		
H x R	68331	1	68331	3.66	
H x R x O	54514	1	54514	2.92	
Error	335809	18	18656		
P x H x R	121471	1	121471	6.16	< .05
P x H x R x O	10372	1	10372	0.53	
Error	354800	18	19711		
S x H x R	1437	1	1437	0.07	
S x H x R x O	131742	1	131742	6.49	< .05
Error	365110	18	20284		
P x S x H x R	63135	1	63135	3.65	
P x S x H x R x O	6106	1	6106	0.35	
Error	311556	18	17309		
D x R	41564	4	10391	0.40	
D x R x O	108811	4	27203	1.05	
Error	1864849	72	25901		
P x D x R	211302	4	52826	4.07	< .01
P x D x R x O	32519	4	8130	0.63	
Error	934307	72	12977		
S x D x R	16673	4	4168	0.18	
S x D x R x O	100559	4	25140	1.07	
Error	1692528	72	23507		
P x S x D x R	63161	4	15790	1.01	
P x S x D x R x O	212367	4	53092	3.40	< .05
Error	1125663	72	15634		
H x D x R	96382	4	24095	0.92	
H x D x R x O	286465	4	71616	2.75	< .05
Error	1875967	72	26055		
P x H x D x R	159079	4	39770	2.35	
P x H x D x R x O	25737	4	6434	0.38	
Error	1218739	72	16927		
S x H x D x R	31391	4	7848	0.42	
S x H x D x R x O	207407	4	51852	2.75	< .05
Error	1355347	72	18824		
P x S x H x D x R	40636	4	10159	0.48	
P x S x H x D x R x O	81055	4	20264	0.96	
Error	1526130	72	21196		

APPENDIX 10 (continued)

C. Summary table of the analysis of variance of mean confidence in correct responses from Experiment 10.

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Order (O)	65.37	1	65.37	1.91	
Error	614.70	18	34.15		
Period (P)	0.02	1	0.02	0.03	
P x O	1.43	1	1.43	2.03	
Error	12.64	18	0.70		
Set (S)	20.79	1	20.79	13.10	< .01
S x O	0.30	1	0.30	0.19	
Error	28.57	18	1.59		
P x S	0.81	1	0.81	1.55	
P x S x O	1.36	1	1.36	2.61	
Error	9.38	18	0.52		
Half (H)	0.01	1	0.01	0.01	
H x O	1.49	1	1.49	3.27	
Error	8.19	18	0.46		
P x H	0.19	1	0.19	0.52	
P x H x O	0.56	1	0.56	1.54	
Error	6.56	18	0.36		
S x H	0.42	1	0.42	1.76	
S x H x O	0.35	1	0.35	1.47	
Error	4.26	18	0.24		
P x S x H	0.07	1	0.07	0.13	
P x S x H x O	0.48	1	0.48	0.93	
Error	9.17	18	0.51		
Discriminability (D)	215.29	4	53.82	30.22	< .001
D x O	1.98	4	0.49	0.28	
Error	128.25	72	1.78		
P x D	0.98	4	0.25	1.11	
P x D x O	0.68	4	0.17	0.77	
Error	15.91	72	0.22		
S x D	5.42	4	1.36	5.41	< .001
S x D x O	0.13	4	0.03	0.13	
Error	18.03	72	0.25		
P x S x D	3.79	4	0.95	5.07	< .01
P x S x D x O	0.83	4	0.21	1.11	
Error	13.43	72	0.19		
H x D	1.00	4	0.25	1.39	
H x D x O	0.89	4	0.22	1.23	
Error	13.01	72	0.18		
P x H x D	0.83	4	0.21	1.09	
P x H x D x O	0.43	4	0.11	0.57	
Error	13.72	72	0.19		
S x H x D	1.19	4	0.30	1.37	
S x H x D x O	0.68	4	0.17	0.79	
Error	15.63	72	0.22		
P x S x H x D	1.04	4	0.26	1.62	
P x S x H x D x O	1.03	4	0.26	1.61	
Error	11.49	72	0.16		

APPENDIX 10 (continued)

C. (continued)

SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F	p
Response (R)	6.63	1	6.63	2.73	
R x O	1.17	1	1.17	0.48	
Error	43.71	18	2.43		
P x R	0.13	1	0.13	0.38	
P x R x O	0.13	1	0.13	0.39	
Error	6.03	18	0.34		
S x R	0.09	1	0.09	0.16	
S x R x O	0.01	1	0.01	0.01	
Error	10.77	18	0.60		
P x S x R	0.04	1	0.04	0.06	
P x S x R x O	0.06	1	0.06	0.09	
Error	12.13	18	0.67		
H x R	0.01	1	0.01	0.01	
H x R x O	0.01	1	0.01	0.01	
Error	6.24	18	0.35		
P x H x R	1.02	1	1.02	3.81	
P x H x R x O	0.03	1	0.03	0.10	
Error	4.82	18	0.27		
S x H x R	0.12	1	0.12	0.48	
S x H x R x O	0.01	1	0.01	0.01	
Error	4.63	18	0.26		
P x S x H x R	0.01	1	0.01	0.01	
P x S x H x R x O	1.89	1	1.89	6.88	< .05
Error	4.99	18	0.27		
D x R	1.35	4	0.35	1.52	
D x R x O	0.65	4	0.16	0.73	
Error	16.00	72	0.22		
P x D x R	0.98	4	0.24	1.67	
P x D x R x O	0.71	4	0.18	1.21	
Error	10.55	72	0.15		
S x D x R	0.65	4	0.16	0.73	
S x D x R x O	0.14	4	0.04	0.16	
Error	16.03	72	0.22		
P x S x D x R	0.08	4	0.02	0.08	
P x S x D x R x O	1.92	4	0.48	2.04	
Error	16.95	72	0.24		
H x D x R	0.67	4	0.17	0.99	
H x D x R x O	0.32	4	0.08	0.47	
Error	12.18	72	0.17		
P x H x D x R	1.65	4	0.41	3.25	< .05
P x H x D x R x O	0.70	4	0.18	1.38	
Error	9.17	72	0.13		
S x H x D x R	0.38	4	0.10	0.58	
S x H x D x R x O	1.06	4	0.27	1.61	
Error	11.86	72	0.16		
P x S x H x D x R	0.53	4	0.13	1.02	
P x S x H x D x R x O	2.10	4	0.52	4.07	< .01
Error	9.29	72	0.13		

BIBLIOGRAPHY

ADAMS, J.K. Behaviour without awareness. Psychological Bulletin, 1957, 54, 383-405.

ADAMS, J.K., & ADAMS, P.A. Realism of confidence judgments. Psychological Review, 1961, 68, 33-45.

ANDERSON, N.H. Scales and Statistics: Parametric and non-parametric. Psychological Bulletin, 1961, 58, 305-316.

ANGELL, F. On judgments of 'like' in discrimination experiments. American Journal of Psychology, 1907, 18, 253-260.

ASCHER, D. A model for confidence judgments in choice tasks. Unpublished thesis submitted for the degree of Ph.D., McMaster University, 1974.

AUDLEY, R.J. A stochastic model for individual choice behavior. Psychological Review, 1960, 67, 1-15.

AUDLEY, R.J. Choosing. Bulletin of the British Psychological Society, 1970, 23, 177-191.

AUDLEY, R.J. Some observations on theories of choice reaction time: Tutorial review. In S. Kornblum (Ed.), Attention and Performance IV. Academic Press: New York, 1973, pp. 509-545.

AUDLEY, R.J., & MERCER, A. The relationship between decision time and relative response frequency in a blue-green discrimination. British Journal of Mathematical and Statistical Psychology, 1968, 21, 183-192.

AUDLEY, R.J., & PIKE, A.R. Some alternative stochastic models of choice. British Journal of Mathematical and Statistical Psychology, 1965, 18, 207-225.

BACH, T.M., CORLETT, J.T., & DICKINSON, J. An information analysis of speed and accuracy. Perceptual and Motor Skills, 1978, 47, 355-359.

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.

BERKSON, J.A. Application of the logistic function to bio-assay. Journal of the American Statistical Association, 1944, 39, 357-365.

BERKSON, J.A. A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. Journal of the American Statistical Association, 1953, 48, 565-599.

BERTELSON, P. Sequential redundancy and speed in a serial two-choice responding task. Quarterly Journal of Experimental Psychology, 1961, 13, 90-102.

BINDRA, D., WILLIAMS, J.A., & WISE, J.S. Judgments of sameness and difference: Experiments on decision time. Science, 1965, 150, 1625-1627.

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.

BIRREN, J.E., & BOTWINICK, J. Speed of response as a function of perceptual difficulty and age. Journal of Gerontology, 1955, 10, 433-436.

BOCK, D.R., & JONES, L.V. The Measurement and Prediction of Judgment and Choice. Holden-Day: San Francisco, 1968.

BONNEL, A.M., & NOIZET, G. Application of signal detection theory to perception of differences in line length. Acta Psychologica, 1979, 43, 1-21.

BORING, E.G. A chart of the psychometric function. American Journal of Psychology, 1917, 28, 465-470.

BOTWINICK, J., BRINLEY, J.F., & ROBBIN, J.S. The interaction effects of perceptual difficulty and stimulus exposure time on age differences in speed and accuracy of response. Gerontologia, 1958, 2, 1-10.

BROADBENT, D.E., & GREGORY, M. Effects of noise and of signal rate upon vigilance analysed by means of decision theory. Human Factors, 1965, 7, 155-162.

BROWN, W. The judgment of difference, with special reference to the doctrine of the threshold, in the case of lifted weights. University of California Publications in Psychology, 1910, 1, 1-71.

BROWN, W. The judgment of very weak sensory stimuli. University of California Publications in Psychology, 1914, 1, 199-268.

BUSH, R.R. Estimation and evaluation. In R.R. Bush, E. Galanter, and R.D. Luce (Eds.), Handbook of Mathematical Psychology, Volume 1. Wiley: New York, 1963, pp. 429-469.

CARLSON, W.R., DRIVER, R.C., & PRESTON, M.G. Judgment times for the method of constant stimuli. Journal of Experimental Psychology, 1934, 17, 113-118.

CARTERETTE, E.C., & COLE, M. Comparison of the receiver-generating characteristics received by the ear and by eye. Journal of the Acoustic Society of America, 1962, 34, 172-178.

CARTERETTE, E.C., FRIEDMAN, M.P., & COSMIDES, R. Reaction-time distributions in the detection of weak signals in noise. Journal of the Acoustic Society of America, 1965, 38, 531-542.

CARTWRIGHT, D. The relation of decision-time to the categories of response. American Journal of Psychology, 1941, 54, 174-196.

CARTWRIGHT, D., & FESTINGER, L. A quantitative theory of decision. Psychological Review, 1943, 50, 595-621.

CATTELL, J.McK. On errors of observation. American Journal of Psychology, 1893, V(3).

CLARK, W.C. The psyche in psychophysics: A sensory-decision theory analysis of the effect of instructions in flicker sensitivity and response bias. Psychological Bulletin, 1966, 65, 358-366.

CLARK, W.C., BROWN, J.C., & RUTSCHMANN, J. Flicker sensitivity and response bias in psychiatric patients and normal subjects. Journal of Abnormal and Social Psychology, 1967, 72, 35-42.

CLARKE, F.R. Confidence ratings, second-choice responses and confusion matrices in intelligibility tests. Journal of the Acoustic Society of America, 1960, 32, 35-46.

COHEN, J., & COHEN, P. Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences. Lawrence Erlbaum: Hillsdale, New Jersey, 1975.

COLQUHOUN, W.P. The effect of "unwanted" signals on performance in a vigilance task. Ergonomics, 1961, 4, 41-51.

COLQUHOUN, W.P. Training for vigilance: A comparison of different techniques. Human Factors, 1966, 8, 7-12.

COLQUHOUN, W.P. Sonar target detection as a decision process. Journal of Applied Psychology, 1967, 51, 187-190.

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.

CORBETT, A. Retrieval dynamics for rote and visual image mnemonics. Journal of Verbal Learning and Verbal Behaviour, 1977, 16, 233-246.

CORBETT, A., & WICKELGREN, W. Semantic memory retrieval: Analysis by speed-accuracy tradeoff functions. Quarterly Journal of Experimental Psychology, 1978, 30, 1-15.

CORSO, J.F. The Experimental Psychology of Sensory Behaviour. Holt, Rinehart and Winston: New York, 1967.

CROSS, D.V., & LANE, H.L. On the discriminative control of concurrent responses: The relations among response frequency, latency and topography in auditory generalisation. Journal of the Experimental Analysis of Behaviour, 1962, 5, 487-496.

CROSSMAN, E.R.F.W. Entropy and choice time: The effect of frequency imbalance on choice response. Quarterly Journal of Experimental Psychology, 1953, 5, 41-51.

CROSSMAN, E.R.F.W. The measurement of discriminability. Quarterly Journal of Experimental Psychology, 1955, 7, 176-195.

CURRY, R.E., NAGEL, D.C., & GAI, E.G. Decision behaviour with changing signal strength. Journal of Mathematical Psychology, 1977, 15, 46-69.

DOSHER, B.A. The retrieval of sentences from memory: A speed-accuracy study. Cognitive Psychology, 1976, 8, 291-310.

DOSHER, B.A. Empirical approaches to information processing: Speed-accuracy tradeoff functions or reaction time - a reply. Acta Psychologica, 1979, 43, 347-359.

EDWARDS, W. Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. Journal of Mathematical Psychology, 1965, 2, 312-329.

EGAN, J.P. Signal Detection Theory and ROC Analysis. Academic Press: New York and London, 1976.

EGAN, J.P., GREENBERG, G.L., & SCHULMAN, A.I. Operating characteristics, signal probability and the method of free response. Journal of the Acoustic Society of America, 1961, 33, 993-1007.

EIJKMAN, C., & VENDRICK, A.J.H. Detection theory applied to the absolute sensitivity of sensory systems. Biophysics Journal, 1963, 3, 65-78.

EMMERICH, D.S., GRAY, J.L., WATSON, C.S., & TANIS, D.C. Response latency, confidence and ROC's in auditory signal detection. Perception and Psychophysics, 1972, 11, 65-72.

FECHNER, G. Elements of Psychophysics. Holt, Rinehart and Winston: New York, 1966. (First published 1860).

FERNBERGER, S.W. On the elimination of the two extreme intensities of the comparison stimuli in the method of constant stimuli. Psychological Review, 1914a, 21, 335-355.

FERNBERGER, S.W. The effect of the attitude of the subject upon the measure of sensitivity. American Journal of Psychology, 1914b, 25, 538-543.

FERNBERGER, S.W. An introspective analysis of the process of comparing. Psychological Monographs, 1919, 26, 1-161.

FERNBERGER, S.W. The use of equality judgments in psychophysical procedures. Psychological Review, 1930, 37, 107-112.

FERNBERGER, S.W. On absolute and relative judgments in lifted weight experiments. American Journal of Psychology, 1931, 43, 560-578.

FERNBERGER, S.W., & IRWIN, F.W. Time relations for the different categories of judgment in the 'absolute method' in psychophysics. American Journal of Psychology, 1932, 44, 505-525.

FERNBERGER, S.W., GLASS, E., HOFFMAN, I., & WILLIG, M. Judgment times of different psychophysical categories. *Journal of Experimental Psychology*, 1934, *17*, 286-293.

FESTINGER, L. Studies in decision: I, Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology*, 1943a, *32*, 291-306.

FESTINGER, L. Studies in decision: II, An empirical test of a quantitative theory of decision. *Journal of Experimental Psychology*, 1943b, *32*, 411-423.

FISCHHOFF, B., SLOVIC, P., & LICHTENSTEIN, S. Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology*, 1977, *3*, 552-564.

FITTS, P.M. Cognitive aspects of information processing: III, Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, *71*, 849-857.

FREEMAN, P.R. Tables of d' and β. Medical Research Council Applied Psychology Unit, Cambridge, 1964.

FULLERTON, G.S., & CATTELL, J.McK. *On the perception of small differences.* University of Pennsylvania Press, 1892, No. 2.

GARRETT, H.E. A study of the relation of accuracy to speed. *Archives of Psychology*, 1922, *56*, 1-104.

GELLER, E.S., & PITZ, G.F. Effects of prediction, probability and run length on choice reaction speed. *Journal of Experimental Psychology*, 1970, *84*, 361-367.

GEORGE, S.S. Attitude in relation to the psychophysical judgment. *American Journal of Psychology*, 1917, *28*, 1-37.

GESCHEIDER, G.A. *Psychophysics: Method and Theory*. Lawrence Erlbaum: New York, 1976.

GESCHEIDER, 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., 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, 10, 413-417.

GLASS, G.V., PECKHAM, P.D., & SANDERS, J.R. Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. Review of Educational Research, 1972, 42, 237-288.

GREEN, D.M., & SWETS, J.A. Signal Detection Theory and Psychophysics. Wiley and Sons: New York, 1966.

GREENBERG, M.G. A modification of Thurstone's law of comparative judgment to accommodate a judgment category of 'equal' or 'no difference'. Psychological Bulletin, 1965, 64, 108-112.

GRICE, G.R. Stimulus intensity and response evocation. Psychological Review, 1968, 75, 359-373.

GRICE, G.R., & HUNTER, J.J. Stimulus intensity effects depend upon the type of experimental design. Psychological Review, 1964, 71, 247-256.

GRICE, G.R., & SPIKER, V.A. Speed-accuracy tradeoff in choice reaction time: Within conditions, between conditions, and between subjects. Perception and Psychophysics, 1979, 26, 118-126.

GUILFORD, J.P. Psychometric Methods. 2nd edition. McGraw-Hill: New York, 1954.

GURALNICK, M.J. Observing responses and decision processes in vigilance. Journal of Experimental Psychology, 1972, 93, 239-244.

HAKE, H.W., & RODWAN, A.S. Perception and recognition. In J.B. Sidowski (Ed.), Experimental Methods and Instrumentation in Psychology. McGraw-Hill: New York, 1966, pp. 331-381.

HALE, D.J. Speed-error tradeoff in a three-choice serial reaction task. Journal of Experimental Psychology, 1969, 31, 428-435.

HARM, O.J., & LAPPIN, J.S. Probability, compatibility, speed, and accuracy. Journal of Experimental Psychology, 1973, 100, 416-418.

HAYDEN, E.A. Memory for lifted weights. American Journal of Psychology, 1906, XVII, 497-521.

HELSON, H. Adaptation-Level Theory. Harper and Row: New York, 1964.

HENMON, V.A.C. The time of perception as a measure of differences in sensation. Archives of Philosophy, Psychology and Scientific Methods, 1906, 8, 1-75.

HENMON, V.A.C. The relation of the time of a judgment to its accuracy. Psychological Review, 1911, 18, 186-201.

HERTZMANN, M. Confidence ratings as an index of difficulty. Journal of Experimental Psychology, 1937, 21, 113-119.

HICK, W.E. On the rate of gain of information. Quarterly Journal of Experimental Psychology, 1952, 4, 11-26.

HOGE, R.D. Confidence in decision as an index of perceived accuracy of information processing. Psychonomic Science, 1970, 18, 351-353.

HORNSBY, P. A test of a model for choice-reaction times. Unpublished thesis submitted for the degree of B.Sc., University of London, 1968.

HOWELL, W.C. An evaluation of subjective probability in a visual discrimination task. Journal of Experimental Psychology, 1967, 75, 479-486.

HOWELL, W.C. Intuitive 'counting' and 'tagging' in memory. Journal of Experimental Psychology, 1970, 85, 210-215.

HOWELL, W.C. Uncertainty from internal and external sources: A clear case of overconfidence. Journal of Experimental Psychology, 1971, 89, 240-243.

HOWELL, W.C., & KREIDLER, D.L. Information processing under contradictory instructional sets. Journal of Experimental Psychology, 1963, 65, 39-46.

HYMAN, R. Stimulus information as a determinant of reaction time. Journal of Experimental Psychology, 1953, 45, 188-196.

INDLIN, Yu. A. The observer as a system with feedback. In H.-G. Geissler and Yu. M. Zabrodin (Eds.), Advances in Psychophysics. VEB Deutscher Verlag der Wissenschaften: Berlin, 1976, pp. 175-193.

IRWIN, F.W., & SMITH, W.A.S. Value, cost, and information as determinants of decision. Journal of Experimental Psychology, 1957, 54, 229-232.

IRWIN, F.W., SMITH, W.A.S., & MAYFIELD, J.F. Tests of two theories of decision in an 'expanded judgment' situation. Journal of Experimental Psychology, 1956, 51, 261-268.

JASTROW, J. A critique of psycho-physic methods. American Journal of Psychology, 1888, 1, 271-309.

JENNINGS, J.R., WOOD, C.C., & LAWRENCE, B.E. Effects of graded doses of alcohol on speed-accuracy tradeoff in choice reaction time. Perception and Psychophysics, 1976, 19, 85-91.

JERISON, H.J., & PICKETT, R.M. Vigilance: The importance of elicited observing rate. Science, 1964, 143, 970-971.

JERISON, H.J., PICKETT, R.M., & STENSON, H.H. The elicited observing rate and decision processes in vigilance. Human Factors, 1965, 7, 107-128.

JERSILD, A. The determinants of confidence. American Journal of Psychology, 1929, 41, 640-642.

JOHNSON, D.M. Confidence and speed in the two-category judgment. Archives of Psychology, 1939, 34, 1-53.

KANTOWITZ, B.H. On the accuracy of speed-accuracy tradeoff. Acta Psychologica, 1978, 42, 79-80.

KATZ, L. A comparison of type II operating characteristics derived from confidence ratings and from latencies. Perception and Psychophysics, 1970, 8, 65-68.

KELLOGG, W.N. The time of judgment in psychometric measures. American Journal of Psychology, 1931, 49, 65-86.

KOHFELD, D.L. Stimulus intensity and adaptation level as determinants of simple reaction time. Journal of Experimental Psychology, 1968, 76, 468-473.

KOPPELL, S. The latency function hypothesis and Pike's multiple observations model for latencies in signal detection. Psychological Review, 1976, 83, 308-309.

KORIAT, A., LICHTENSTEIN, S., & FISCHHOFF, B. Reasons for confidence. Journal of Experimental Psychology, 1980, 6, 107-118.

LA BERGE, D. Generalisation gradients in a discrimination situation. Journal of Experimental Psychology, 1961, 62, 88-94.

LA BERGE, D. A recruitment theory of simple behaviour. Psychometrika, 1962, 27, 375-396.

LAMB, J., & KAUFMAN, H. Information transmission with unequally likely alternatives. Perceptual and Motor Skills, 1965, 21, 255-259.

LAMING, D.R.J. A statistical test of a prediction from information theory in a card-sorting situation. Quarterly Journal of Experimental Psychology, 1962, 14, 38-48.

LAMING, D.R.J. Information Theory of Choice Reaction Times. Academic Press: New York and London, 1968.

LAMING, D.R.J. Subjective probability in choice-reaction experiments. Journal of Mathematical Psychology, 1969, 6, 81-120.

LAPPIN, J.S., & DISCH, K. The latency operating characteristic: I. Effects of stimulus probability on choice reaction time. Journal of Experimental Psychology, 1972a, 92, 419-427.

LAPPIN, J.S., & DISCH, K. The latency operating characteristic: II. Effects of visual stimulus intensity on choice reaction time. Journal of Experimental Psychology, 1972b, 93, 367-372.

LAPPIN, J.S., & DISCH, K. Latency operating characteristic: III. Temporal uncertainty effects. Journal of Experimental Psychology, 1973, 98, 279-285.

LEMMON, V.W. The relation of reaction time to measures of intelligence, memory, and learning. Archives of Psychology, 1927, No. 94.

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.

LIEBLICH, I., & LIEBLICH, A. Effects of different payoff matrices on arithmetical tasks: An attempt to produce "rationality". Perceptual and Motor Skills, 1969a, 29, 467-473.

LIEBLICH, A., & LIEBLICH, I. Arithmetical estimation under conditions of different payoff matrices. Psychonomic Science, 1969b, 14, 87.

LINK, S.W. Applying RT deadlines to discrimination reaction time. *Psychonomic Science*, 1971, *25*, 355-358.

LINK, S.W. The relative judgment theory of two choice response time. *Journal of Mathematical Psychology*, 1975, *12*, 114-135.

LINK, S.W. The relative judgment theory of the psychometric function. In J. Requin (Ed.), *Attention and Performance VII*. Lawrence Erlbaum: Hillsdale, New Jersey, 1978a, pp. 619-630.

LINK, S.W. The relative judgment theory analysis of response time deadline experiments. In N.J. Castellan Jr. and F. Restle (Eds.), *Cognitive Theory III*. Lawrence Erlbaum: Hillsdale, New Jersey, 1978b, pp. 117-138.

LINK, S.W. Improvements on a new model for choice reaction time. *Perception and Psychophysics*, 1979, *25*, 443-446.

LINK, S.W., & HEATH, R.A. A sequential theory of psychological discrimination. *Psychometrika*, 1975, *40*, 77-105.

LINK, S.W., & TINDALL, A.D. Speed and accuracy in comparative judgments of line length. *Perception and Psychophysics*, 1971, *9*, 284-288.

LITTLE, K.B., & LINTZ, L.M. Information and certainty. *Journal of Experimental Psychology*, 1965, *70*, 428-432.

LIVELY, B.L. Speed/accuracy tradeoff and practice as determinants of stage durations in a memory-search task. *Journal of Experimental Psychology*, 1972, *96*, 97-103.

LOEB, M., & BINFORD, J.R. Vigilance for auditory intensity changes as a function of preliminary feedback and confidence level. *Human Factors*, 1964, *7*, 445-458.

LOEB, M., & BINFORD, J.R. Variation in performance on auditory and visual monitoring tasks as a function of signal and stimulus frequencies. *Perception and Psychophysics*, 1968, *4*, 361-367.

LUCE, R.D. *Individual Choice Behaviour*. Wiley: New York, 1959.

LUND, F.H. The criteria of confidence. *American Journal of Psychology*, 1926, *37*, 372-381.

LYONS, J.J., & BRIGGS, G.E. Speed-accuracy trade-off with different types of stimuli. Journal of Experimental Psychology, 1971, 91, 115-119.

MACKWORTH, J.F. Deterioration of signal detectability during a vigilance task as a function of background event rate. Psychonomic Science, 1965, 3, 421-422.

MARTIN, L.J., & MULLER, G.E. Zur Analyseder Unterschiedsempfindlichkeit. Leipzig, 1899.

McGILL, W.J. Stochastic latency mechanisms. In R.D. Luce, R.R. Bush, and E. Galanter (Eds.), Handbook of Mathematical Psychology, Volume 1. Wiley: New York, 1963, pp. 309-360.

McNICOL, D. A Primer of Signal Detection Theory. Allen and Unwin: London, 1972.

MORGAN, B.B., & ALLUISI, E.A. Effects of discriminability and irrelevant information on absolute judgment. Perception and Psychophysics, 1967, 2, 54-58.

MURRAY, H.G. Stimulus intensity and reaction time: Evaluation of a decision theory model. Journal of Experimental Psychology, 1970, 84, 383-391.

NICKERSON, R.S. "Same-different" response times: A model and a preliminary test. In W.G. Koster (Ed.), Attention and Performance II. New Holland Publishing Company: Amsterdam, 1969, pp. 257-275.

NICKERSON, R.S. "Same-different" response times. A further test of a "counter and clock" model. Acta Psychologica, 1971, 35, 112-127.

OGILVIE, J.C., & CREELMAN, C.L. Maximum likelihood estimation of receiver operating characteristic curve parameters. Journal of Mathematical Psychology, 1968, 5, 377-391.

OLSON, C.L., & OGILVIE, J.C. The method of constant stimuli with two or more categories of response. Journal of Mathematical Psychology, 1972, 9, 320-338.

OSGOOD, C.E. Method and Theory in Experimental Psychology. Oxford University Press: New York, 1953.

PACHELLA, R.G. The interpretation of reaction time in information processing research. In B.H. Kantowitz (Ed.), Human Information Processing: Tutorials in Performance and Cognition. Lawrence Erlbaum: Hillsdale, New Jersey, 1974, pp. 41-82.

PACHELLA, R.G., & FISHER, D.F. Effect of stimulus degradation and similarity in the tradeoff between speed and accuracy in absolute judgments. Journal of Experimental Psychology, 1969, 81, 7-9.

PACHELLA, R.G., & PEW, R.W. Speed-accuracy tradeoff in reaction time: Effect of discrete criterion times. Journal of Experimental Psychology, 1968, 76, 19-24.

PACHELLA, R.G., FISHER, D.F., & KARSH, R. Absolute judgments in speeded tasks: Quantification of the tradeoff between speed and accuracy. Psychonomic Science, 1968, 12, 225-226.

PEARSON, E.S., & HARTLEY, H.D. Biometrika tables for statistics Vol. 1. (3rd ed.). Cambridge University Press: London, 1976.

PETERS, W. Uber Ahnlichkeitsassoziation. 1910.

PEW, R.W. The speed-accuracy operating characteristic. In W.G. Koster (Ed.), Attention and Performance II. New Holland Publishing Company: Amsterdam, 1969, pp. 16-26.

PHILIP, B.R. The relationship between speed and accuracy in a motor task. Journal of Experimental Psychology, 1936, 19, 24-50.

PHILIP, B.R. The relationship of exposure time and accuracy in a perceptual task. Journal of Experimental Psychology, 1947, 37, 178-186.

PICKETT, R.M. The perception of a visual texture. Journal of Experimental Psychology, 1964, 68, 13-20.

PICKETT, R.M. Response latency in a pattern perception situation. Acta Psychologica, 1967, 27, 160-169.

PICKETT, R.M. The visual perception of random line segment texture. Paper read at the Ninth Annual Meeting of the Psychonomic Society, 1968.

PIERCE, C.S., & JASTROW, J. On small differences of sensation. Memoirs of the National Academy of Sciences, 1885, 3, 74-83.

PIERREL, R., & MURRAY, C.S. Some relationships between comparative judgment, confidence and decision-time in weight lifting. American Journal of Psychology, 1963, 76, 28-38.

PIKE, A.R. Latency and relative frequency of response in psychophysical discrimination. British Journal of Mathematical and Statistical Psychology, 1968, 21, 161-182.

PIKE, A.R. The latencies of correct and incorrect responses in discrimination and detection tasks: Their interpretation in terms of a model based on simple counting. Perception and Psychophysics, 1971, 9, 455-460.

PIKE, A.R. Response latency models for signal detection. Psychological Review, 1973, 80, 53-68.

PIKE, A.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, A.R., & DALGLEISH, L. Latency-probability curves for sequential decision models: A comment on Weatherburn. Psychological Bulletin, 1982, 91, 384-388.

PIKE, A.R., & RYDER, P. Response latencies in the yes/no detection task: An assessment of two basic models. Perception and Psychophysics, 1973, 13, 224-232.

PIKE, A.R., & RYDER, P. Multiple observations and latency functions: A further note on response latency in signal detection. Psychological Review, 1978, 85, 48-52.

PIKE, A.R., McFARLAND, K., & DALGLEISH, L. Speed-accuracy tradeoff models for auditory detection with deadlines. Acta Psychologica, 1974, 38, 379-399.

POLLACK, I., & DECKER, L.R. Confidence ratings, message reception, and the receiver operating characteristics. Journal of the Acoustic Society of America, 1958, 30, 286-292.

POULTON, E.C. Unwanted range effects from using within-subject experimental design. Psychological Bulletin, 1973, 80, 113-121.

POULTON, E.C. Range effects in experiments on people. American Journal of Psychology, 1975, 88, 3-32.

POULTON, E.C. Models for biases in judging sensory magnitude. Psychological Bulletin, 1979, 86, 777-803.

RABBITT, P.M.A. Times for the analysis of stimuli and for the selection of responses. In A. Summerfield (Ed.), Cognitive Psychology. British Medical Bulletin, 1971, 27, 259-265.

RATCLIFF, R. Group reaction time distributions and an analysis of distribution statistics. Psychological Bulletin, 1979, 86, 446-461.

REED, A.V. Speed-accuracy tradeoff in recognition memory. Science, 1973, 181, 574-576.

REED, A.V. List length and the time course of recognition in immediate memory. Memory and Cognition, 1976, 4, 16-30.

REMINGTON, R.J. Analysis of sequential effects for a four-choice reaction time experiment. Journal of Psychology, 1971, 77, 17-27.

SANDERS, A.F., & TER LINDEN, W. Decision making during paced arrival of probabilistic information. In A.F. Sanders (Ed.), Attention and Performance II. Acta Psychologica, 1967, 27, 170-177.

SCHMITT, J.C., & SCHEIRER, C.J. Empirical approaches to information processing: Speed-accuracy tradeoff functions or reaction time. Acta Psychologica, 1977, 41, 321-325.

SCHOUTEN, J.F., & BEKKER, J.A.M. Reaction time and accuracy. In A.F. Sanders (Ed.), Attention and Performance II. Acta Psychologica, 1967, 27, 143-153.

SEKULER, R.W. Signal detection, choice response times, and visual backward masking. Canadian Journal of Psychology, 1965, 19, 118-132.

SEWARD, G. Recognition time as a measure of confidence. Archives of Psychology, 1928, No. 99.

STEVENS, S.S. Mathematics, measurement and psychophysics. In S.S. Stevens (Ed.), Handbook of Experimental Psychology. Wiley: New York, 1951, pp. 1-49.

STONE, M. Models for choice reaction time. Psychometrika, 1960, 25, 251-260.

SWANSON, J.M., & BRIGGS, G.E. Information processing as a function of speed versus accuracy. Journal of Experimental Psychology, 1969, 81, 223-229.

SWENSSON, R.G. The elusive tradeoff: Speed versus accuracy in visual discrimination tasks. Perception and Psychophysics, 1972a, 12, 16-32.

SWENSSON, R.G. Tradeoff bias and efficiency effects in serial choice reactions. Journal of Experimental Psychology, 1972b, 95, 397-407.

SWETS, J.A. Signal detection theory applied to vigilance. In R. Mackie (Ed.), Vigilance: Theory, Operational Performance, and Physiological Correlates. Plenum Press: New York, 1977, pp. 705-718.

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 C. Cherry (Ed.), Information Theory: Proceedings of the Fourth London Symposium. Butterworths: London, 1961, pp. 177-195.

SWETS, J.A., TANNER, W.P.Jr., & BIRDSALL, T.G. Decision processes in perception. Psychological Review, 1961, 68, 301-340.

TANNER, W.P.Jr. Theory of recognition. Journal of the Acoustic Society of America, 1956, 28, 882-888.

TANNER, W.P.Jr. What is masking? Journal of the Acoustic Society of America, 1958, 30, 919-921.

TANNER, W.P.Jr., & SWETS, J.A. A decision-making theory of visual detection. Psychological Review, 1954, 61, 401-409.

TAYLOR, M.M. Detectability measures in vigilance: Comment on a paper by Wiener, Poock and Steele. Perceptual and Motor Skills, 1965, 20, 1217-1221.

TAYLOR, M.M., LINDSAY, P.H., & FORBES, S.M. Quantification of shared capacity processing in auditory and visual discrimination. In A.F. Sanders (Ed.), Attention and Performance II. Acta Psychologica, 1967, 27, 223-229.

THEIOS, J. The Locus of Cognition. Paper presented at the Thirteenth Annual Meeting of the Psychonomic Society, 1972.

THEIOS, J., SMITH, P.G., HAVILAND, S.E., TRAUPMAN, J., & MOY, M.C. Memory scanning as a serial self-terminating process. Journal of Experimental Psychology, 1973, 97, 323-336.

THOMAS, E.A.C., & MYERS, J.L. Implications of latency data for threshold and nonthreshold models of signal detection. Journal of Mathematical Psychology, 1972, 9, 253-258.

THOMSON, G.H. A new point of view in the interpretation of threshold measurements in psychophysics. Psychological Review, 1920, 27, 300-307.

THURSTONE, L.L. Psychophysical analysis. American Journal of Psychology, 1927a, 38, 368-389.

THURSTONE, L.L. A law of comparative judgment. Psychological Review, 1927b, 34, 273-286.

THURSTONE, L.L. The phi-gamma hypothesis. Journal of Experimental Psychology, 1928, 11, 293-305.

THURSTONE, L.L. Ability, motivation and speed. Psychometrika, 1937, 2, 249-254.

THURSTONE, L.L. Psychophysical methods. In T.G. Andrews (Ed.), Methods of Psychology. Wiley: New York, 1948, pp. 124-157.

TREISMAN, M., & WATTS, T.R. Relation between signal detectability theory and the traditional procedures for measuring sensory threshold: Estimating d' from results given by the method of constant stimuli. Psychological Bulletin, 1966, 66, 438-454.

TROW, W.C. The psychology of confidence. Archives of Psychology, 1923, 67, 1-47.

URBAN, F.M. The method of constant stimuli and its generalisations. Psychological Review, 1910, 17, 229-259.

VICKERS, D. Theories and experiments on visual discrimination and the perception of visual depth. Unpublished Ph.D. Thesis, Cambridge University, 1967.

VICKERS, D. Evidence for an accumulator model of psychophysical discrimination. In A.T. Welford and L. Houssiadis (Eds.), Current Problems in Perception. Ergonomics, 1970, 13, 37-58.

VICKERS, D. A cyclic decision model of perceptual organisation. Perception, 1972, 1, 31-48.

VICKERS, D. Where Angell feared to tread: Response time and frequency in three-category discrimination. In P.M.A. Rabbitt and S. Dornic (Eds.), Attention and Performance V. Academic Press: London, 1975, pp. 455-469.

VICKERS, D. An adaptive module for simple judgments. In J. Requin (Ed.), Attention and Performance VII. Lawrence Erlbaum: Hillsdale, New Jersey, 1978, pp. 599-618.

VICKERS, D. Decision Processes in Visual Perception. Wiley: New York and London, 1979.

VICKERS, D. Unpublished simulations of the accumulator model. The University of Adelaide, 1981.

VICKERS, D., & LEARY, J. Criterion control in signal detection. Human Factors, 1983, 25, 283-296.

VICKERS, D., CAUDREY, D., & WILLSON, R.J. Discriminating between the frequency of occurrence of two alternative events. Acta Psychologica, 1971, 35, 151-172.

VICKERS, D., LEARY, J., & BARNES, P. Adaptation to decreasing signal probability. In R.R. Mackie (Ed.), Vigilance: Theory, Operational Performance, and Physiological Correlates. Plenum Press: New York, 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.

VOLKMANN, J. The relation of time of judgment to certainty of judgment. Psychological Bulletin, 1934, 31, 672-673.

WEATHERBURN, D. Latency-probability functions as bases for evaluating competing accounts of the sensory decision process. Psychological Bulletin, 1978, 85, 1344-1347.

WEATHERBURN, D., & GRAYSON, D. Latency-probability functions: A reply to Pike and Dalgleish. Psychological Bulletin, 1982, 91, 389-392.

WELFORD, A.T. Fundamentals of Skill. Methuen: London, 1968.

WELFORD, A.T. What is the basis of choice reaction-time? Ergonomics, 1971, 14, 679-693.

WELFORD, A.T. Attention, strategy and reaction time: A tentative metric. In S. Kornblum (Ed.), Attention and Performance IV. Academic Press: London and New York, 1973, pp. 37-53.

WELFORD, A.T. Display layout, strategy and reaction time: Tests of a model. In P.M.A. Rabbitt and S. Dornic (Eds.), Attention and Performance V. Academic Press: London and New York, 1975, pp. 470-484.

WICKELGREN, W.A. Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica, 1977, 41, 67-85.

WILDING, J.M. Effects of stimulus discriminability on the latency distribution of identification responses. Acta Psychologica, 1974, 38, 483-500.

WILDING, J.M. An investigation of possible response effects on the latency distribution of identification responses. Acta Psychologica, 1978, 42, 231-251.

WILLIAMSON, G.F. Individual differences in belief measured and expressed by degrees of confidence. Journal of Philosophy, Psychology and Scientific Method, 1915, 12, 127-137.

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.

WINER, B.J. Statistical Principles in Experimental Design. McGraw-Hill Kogakusha: London, 1972.

WOLLEN, K.A. Relationship between choice time and frequency during discrimination training and generalization tests. Journal of Experimental Psychology, 1963, 66, 474-484.

WOOD, C.C., & JENNINGS, J.R. Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. Perception and Psychophysics, 1976, 19, 92-101.

WOODWORTH, R., & SCHLOSBERG, H. Experimental Psychology. Holt: New York, 1954.