A detection theory investigation of sensitivity and bias in recognition memory.

A thesis presented in partial
fulfillment of the requirements for the degree of
Master of Arts in Psychology
at Massey University, Albany Campus

Jane M. Burgess
1999

Abstract

Within Signal Detection Theory (SDT) it is accepted that the measures of sensitivity and bias are independent of each other. However, the independence of bias and sensitivity with respect to an individual's behaviour is uncertain. In the current study two experiments were completed to investigate this question. In a Yes/No recognition memory task for words, eight participants each completed 27 blocks of 120 trials, presented and scored on a computer. Nine blocks were completed in Experiment 1 and 18 in Experiment 2. In Experiment 1 sensitivity was altered by means of changes in word imagery levels. Measures of sensitivity and response bias were obtained when participants were tested with either high, medium, or low imagery words. In Experiment 2 bias was manipulated by artificially weighting the consequences of correct and incorrect responses. Analysis of the results was undertaken using both parametric and nonparamertric SDT measures of sensitivity and bias. Analyses of variance showed that there were no statistically significant relationships between imagery level (an indirect measure of sensitivity) and response bias. However, a correlational analysis between the individual sensitivity measures and response bias indicated that, when there were no external biasing factors, response bias became less pronounced response bias as sensitivity increased. The study also indicated that participants' natural response biases tended to be conservative.

Acknowledgments

First and foremost, I would like to thank my supervisor, Jennifer Stillman, whose support, understanding, knowledge and patience has enabled me to tackle and complete this thesis.

I would also like to thank those students and friends who volunteered to participate in this research. Your efforts and endurance are much appreciated, and without you this study would not have been possible.

Finally, I would like to, separately, extend my most sincere thanks to my family and friends who have supported me and who have been more than understanding - it was greatly appreciated.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

Rationale

The purpose of the current research was to assess whether response bias is affected by sensitivity in a word recognition task. This particular issue is of interest in relation to the work of Snodgrass and Corwin (1988) who concluded from their research that different groups of people can be distinguished by the types of bias they adopt. More specifically, Snodgrass and Corwin claimed that amnesic individuals tend to be more liberal in their criterion placement than normal individuals. However, in their study the task difficulty was not equal for the two groups; rather the level of difficulty was higher for the amnesic participants. While it is accepted that Signal Detection Theory (SDT, the approach used here) has the ability to measure sensitivity and bias as two independent variables, it is not known with certainty whether individuals' behaviour, with respect to their response tendencies, is independent of sensitivity. If it is not, Snodgrass and Corwins' conclusions maybe somewhat premature because bias comparisons were made across different sensitivities. The purpose of this study is to examine whether sensitivity and bias are independent with respect to a participant's behaviour.

Signal Detection Theory

Psychophysics is the study of the relationship between physical stimuli and the psychological experiences to which they give rise. The term is also applied to the techniques used to quantify people's perceptions of physical realities (Hannay, 1986). The theory of psychophysics was originally developed in Germany in the late nineteenth century. Its primary developers were Wilhelm Wundt (1832 - 1920), and Gustav Fechner (1801 - 1887) (Carlson, 1993).

In psychophysical investigations techniques to obtain objective measures are needed. Some of these are applied within the theoretical framework of SDT. SDT is a step away from classical psychophysics and was applied to human judgment by Tanner and Swets (1954) and Swets, Tanner, and Birdsall (1961, cited in Swets, 1986). Based on Wald's (1950, cited in Swets, 1986) statistical decision theory, SDT developed during World War II through efforts to enhance the monitoring of new technology such as radar. The difference between this theory and other theories is firstly, SDT assumes that individuals are constantly receiving sensory input consisting of both a stimulus, often referred to as the 'signal', and extraneous input from within the nervous system or from the environment, referred to as 'noise' (Hannay, 1986). SDT was considered a marked improvement on older methods used to measure sensory thresholds (method of limits and method of constant stimuli) because of its ability to measure sensitivity and response bias as two independent factors, thus giving more accurate estimations of performance (Curry, Nagel, & Gai, 1977; Green & Swets, 1966; Hilgendorf & Irving, 1978; Singh & Churchill, 1986; Stubbs, 1976). However, some have argued quite convincingly that the aforementioned classical methods could also obtain these independent measures (Gardner, 1997).

SDT is a method used by experimenters to assess an individual's response to stimuli within the environment. The experimenter can do this by examining the participants sensitivity and response bias, which can be depicted numerically or graphically (in the form of a receiver operating characteristic curve - ROC curve)[1] (Green & Swets, 1966). For example, SDT has been applied to human memory in such a way that effects of memory-facilitating drugs on patients can be assessed. For example, Bartus, Dean, Beer, & Lippa (1982) used SDT to establish the effects particular drugs, aimed at specific neurotransmitters, had on memory, and to assess support for the cholinergic hypothesis of geriatric memory dysfunction. Thus, SDT has been used to establish the effectiveness of drugs and also to help pinpoint the contributing factors to diseases such as Alzheimer's. This would not be possible using a measure such as percent correct, that confounds sensitivity and response bias.

Several psychophysical tasks can be used to provide data for a detection theory analysis. The so-called 'forced choice' method is usually employed to obtain data for the investigation of recognition memory. The following discussion relates to this type of task wherein two responses are available to the observer: one response when the stimulus to be detected is judged to be present and another when it is judged to be absent. For example, in a Yes/No recognition memory task for words, respondents say 'yes' when they think a test word was in a previously seen list, or 'no' when they think that it was not included in the list.

---

[1] The terms sensitivity and response bias will be discussed in more detail later, with reference to how test scores are derived and what they imply. The argument concerning whether these two aspects of behaviour are truly independent will also be addressed,

In a detection task four types of outcome are possible. The response can be categorised either as a *hit,* a *miss,* a *correct rejection* or a *false alarm.* A hit occurs when a stimulus was present and the participant states there was a stimulus. A miss occurs when a stimulus was present and the participant states that there was no stimulus. A correct rejection occurs when the participant responds that there was no stimulus, and no stimulus was presented, and finally a false alarm occurs when the participant responds that there was a stimulus when in fact there was no stimulus. These outcomes are usually displayed in a two-by-two matrix as follows (Green & Swets, 1966):

		Response: Yes	Response: No
Stimulus	Present	Hit	Miss
	Absent	False Alarm	Correct Rejection

Figure 1. Matrix displaying the four possible responses that can be made by a participant in a Yes/No SDT task.

A single pair of scores, either hits and false alarms, or misses and correct rejections, completely describe the outcome of a series of trials. Usually the hit and false alarm pairs are chosen. Each square in the matrix as displayed in Figure 1 can be assigned a conditional probability. For example, 'hits' can be expressed as P(Y|y), where 'Y' is the response and 'y' is the stimulus. This represents the probability that the participant responded yes, given that the stimulus was present (McNicol, 1972). The conditional probabilities for all the response categories are displayed in Figure 2 below. It is from the proportion of hits and the proportion of false alarms that d′ (an SDT measure of sensitivity) can be derived (Green & Swets, 1966).

<table>
<tr><td></td><td></td><td colspan="2">Response</td></tr>
<tr><td></td><td></td><td>Yes</td><td>No</td></tr>
<tr><td rowspan="2">Stimulus</td><td>Present</td><td>Hit
P(Y|y)</td><td>Miss
P(N|y)</td></tr>
<tr><td>Absent</td><td>False
Alarm
P(S|n)</td><td>Correct
Rejection
P(N|n)</td></tr>
</table>

Figure 2. Matrix displaying conditional probabilities of the four possible responses.

Recognition Memory

Memory is the process of storing information in the brain. There are three main steps in remembering information. Firstly, the information must be encoded, and secondly it must be stored. Finally, for the information to be remembered, it must be retrieved. Two basic forms of memory are recall and recognition. Recall involves remembering information without specific cues. Recognition memory concerns the ability to pick a previously presented stimulus from a set of novel stimuli (Kaschak & Charnetski, 1998; Snodgrass & Corwin, 1988). The most common experimental procedure consists of presenting a series of stimuli, in this case a list of words referred to as the memory list (old words), followed by the presentation of a second list comprised of both 'old' and 'new' (distractor) words (Brown, Lewist, & Monk, 1977; Butters, Wolfe, Marlone, Granholm, & Cermak, 1985; Estes & Maddox, 1995; Noldy, Stelmack, & Campbell, 1990). The task is to correctly identify the 'old' and 'new' items. A classic example of a test of recognition memory is the multiple choice test.

How an individual carries out the process of recognition is a topic of much debate. Most recent research indicates that recognition is an explicit conscious process within long-term memory (Reed & Squire, 1997) and is based on the experience of familiarity (Brown, 1976; Hirshman, 1995; Hirshman & Henzler, 1998).

Within detection theory it is assumed that recognition is based upon the familiarity of an item. It is also assumed that familiarity is a normally distributed variable, with the mean familiarity of old items being higher than that of new items. It is claimed that a participant will set a criterion along the familiarity dimension, and if an item has greater familiarity than the criterion it will result in a response of 'old', whereas an item with a familiarity below the criterion will result in a response of 'new' (Shiffrin & Steyvers, 1997).

When SDT is applied to memory experiments, certain changes to the terms that are used are appropriate. Firstly, signal-plus-noise refers to the familiarity distribution associated with old words, and noise refers to the familiarity distribution associated with new words. The term ROC can also be replaced with the term MOC which stands for the Memory Operating Characteristic (Banks, 1970; Wicklegren & Norman, 1966). Recently, the familiarity principle has been expanded into the remember-know paradigm where SDT has become the main rival of a dual process model. In this situation, rather than having just one criterion, there are two which have different familiarity values. Anything above the criterion with the higher value is considered a 'know' stimulus and anything beneath the lower value criterion considered a 'remember' stimulus. Stimuli falling in between are responded to as a 'know'. While some of the details are debatable (Gardiner & Gregg, 1997) most research has found support for this theory (Donaldson, 1996; Hirshman, 1995; Hirshman & Henzler, 1998). Due to the relative newness of the remember-know paradigm and the level of difficulty involved in implementing it experimentally, this approach was not adopted in the current study, which was, in any case, intended to replicate the paradigm adopted by Snodgrass and Corwin (1988).

Remembrance depends on how well something was encoded. To investigate whether response bias is independent of sensitivity in a recognition task for words, the level of difficulty, and thus sensitivity, was manipulated by choosing words categorised as high or low on the dimensions of imagibility. Sensitivity is discussed below, as is imagery.

Sensitivity

Sensitivity is a measure of how well a particular individual can discriminate between two events, namely, noise (new words) and signal plus noise (old words). For both response bias and sensitivity, there are several alternatives measures that can be used. For example, common measures of sensitivity include the area under the ROC, A', d' and d'_e (Macmillan & Creelman, 1996).

The receiver operating characteristic (ROC) curve is a nonparametric measure for sensitivity. This name originally stemmed from its use in radar and signal detection, and from its use in engineering (Luce, 1963, cited in Macmillan & Creelman, 1991). Since then it has been renamed the relative operating characteristic (Swets, 1986). As mentioned earlier, when applied to memory experiments, the term MOC is usually employed (Banks, 1970; Wicklegren & Norman, 1966). Another term that is often used is the isosensitivity curve, indicating that all the points on the curve have the same sensitivity (Macmillan & Creelman, 1991).

In the typical SDT experiment, the results can be summarised in the two-by-two matrix first introduced in Figures 1 and 2. The ROC curve is obtained by fitting a theoretical function to a number of hit and false alarm pairs. The best fitting theoretical curve is fitted to a number of data points obtained either from a rating experiment or from several pairs of hits and false alarms obtained either by changing signal probability or by manipulating the payoffs of correct and incorrect responses, as was done in this experiment.

The current experiment used single-point estimates of the ROC because when bias is the variable of intent, most investigators use a binary task, as did Snodgrass and Corwin (1988).

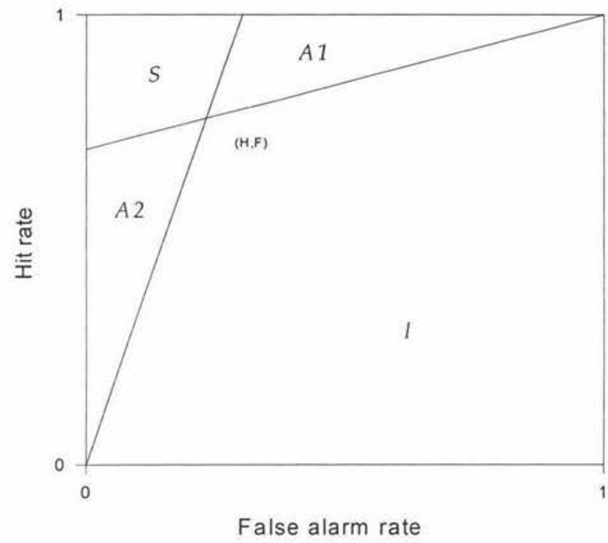

Figure 3. Geometric derivation of A′ and B″ for a single pair (hit and false-alarm) in ROC space.

In 1964, Green (cited in Macmillan & Creelman, 1996) published his finding that the area under the Yes/No ROC curve was a useful measure of sensitivity and was assumption free. This is so because, by Green's theorem, the area under the ROC is equal to the predicted proportion correct by an unbiased observer in a two alternative forced choice (2AFC) experiment. For further discussion see Macmillan and Creelman (1990, 1991). Stemming from this finding, Pollack and Norman's (1964) measure, A′, was developed. A′ is an estimate of the area under the ROC and is derived geometrically. In theory it is the ideal measure because it can be based on a one point estimate, and it is assumption free. With reference to Figure 3, 'S' represents the area which cannot lie under the true ROC, and 'I' indicates the area that does lie under the curve. A1 and A2 are the areas that may or may not lie under the true ROC in whole or in part. Thus, A′ is found by following the geometric equation:

$$A' = I + .5(A_1 - A_2).$$

(1)

A' is the most commonly used nonparamteric measure, and it is usually paired with the bias measure B'', discussed later (Macmillan & Creelman, 1996).

The most commonly used parametric index is d', which is the measure principally used by Snodgrass and Corwin (1988). d' is the normalised distance between the means of the hypothetical sensory distributions associated with the noise and signal-plus-noise which, in the simplest model, are both normal, and are of equal variance. d' is gained by the use of the inverse normal distribution function, otherwise known as *z*-scores which can be obtained in most statistical text books (Wild & Seber, 1994). Given normal-normal equal variance assumptions for the underlying distributions, d' can be found using the equation below:

$$d' = z(H) - z(F).$$

(2)

Thus, sensitivity is specified as the distance between the two means ($x_o - x_n$, where the subscripts 'o' and 'n' refer to the old and new words respectively). When sensitivity is low the two means will be close together, and when sensitivity is high the means will be far apart.

It is not uncommon for the experimenter to attempt to manipulate bias. The advantage of using SDT is that the measures of bias and sensitivity are independent of each other.

Therefore, the experimenter can manipulate response bias without affecting the measure of sensitivity and, presumably vice versa (Macmillan & Creelman, 1990).

To illustrate this point consider the responding of a hypothetical participant. In the first block of trials the participant obtains a hit rate of .60 and a false alarm rate of .20. Then, in the second block of trials the experimenter alters the payoff for correct and incorrect responses promoting a shift in the participant's criterion favouring conservative responses. The participant becomes more conservative, producing a hit rate of .35 and a false alarm rate of .07. Finally, the payoff matrix is altered again so that the participant adopts a more liberal criterion, resulting in a hit rate of .80 and a false alarm rate of .40. From these data d′ can be calculated as follows (making use of Equation 2):

(a) $z(.60) - z(.20)$

$(.253) - (-.842) = 1.095$

(b) $z(.35) - z(.07)$

$(-.385) - (-1.476) = 1.091$

(c) $z(.80) - z(.40)$

$(.842) - (-.253) = 1.095$

This illustrates constant sensitivity although the criterion has shifted as a result of changes in the weighting of the payoff matrix (McNicol, 1972).

It is important to understand the discrimination process that lies behind the measurement of decision space. This can be achieved by investigating internal representations, or decision space. As mentioned earlier, SDT assumes that sensory experiences arising from a constant stimulus are not fixed, but vary. In the current research, it is assumed that the familiarity of words is normally distributed. Thus, both the 'old' and 'new' words will have a normal distribution of familiarity, each with a different mean, with the previously seen words having the higher mean.

In the simplest model, when measuring sensitivity, it is assumed that the variances (or standard deviations) of the distributions of interest are equal. For example, in the current experiment it is assumed that the variance of familiarity for the old words will be the same as the variance for the new words. When this is the case, the slope of the ROC, which forms a straight line on bivariate normal axes, will equal one. On occasion, however, the slope does not equal one, and this indicates that in some situations the variance for old words may be greater or lesser than that for new words. When this occurs, there are two common alternative measures of sensitivity. The first, Δm, was proposed by Green and Swets (1966), and is obtained using the equation below.

$$\Delta m = (x_o - x_n)/ \sigma_n. \tag{3}$$

That is, Δm (change in the mean) is the mean of the old words (x_o) minus the mean of the new words (x_n), divided by standard deviation of the old words (σ_n).

However, later criticisms by Egan (1975, cited in Irwin, Hautus, & Stillman, 1992) suggest that his alternative, d'_e, is the best alternative when faced with normal-normal, unequal variance distributions. Egan's suggestion can be obtained by following the equation below:

$$d'_{e} = x_o - x_n /(0.5(\sigma_o - \sigma_n)). \tag{4}$$

That is, the distance between the means of the distribution for the old and the new words, divided by the average value of the two standard deviations. This index of sensitivity is useful when the variances are unequal, and will reduce to d′ where the variances are, in fact, equal. By using this equation, a measurement is produce which represents a compromise between the two standard deviations. The usual assumption is that each is an estimate of a single quantity, the common standard deviation.

Measures of Response Bias

Response bias represents an individual's tendency to opt for one or the other of the available response alternatives when in situations of uncertainty. Thus, in the current experiment, when individuals are unsure whether they have seen a word before, those with liberal bias will be more inclined to respond 'old', and those with a conservative bias will be more inclined to respond 'new'.

There are several parametric and nonparametric measures of response bias (See, Warm, Dember, & Howe, 1997). The function of the measure of response bias is to locate the criterion which represents the deciding line between experiences resulting in different response alternatives. The setting of a criterion leads to the use of one response alternative for stimuli above a set signal strength, and the other response alternative for those beneath it. The criterion is considered to be a dimensionless quantity. An example criterion depicting no bias is represented in Figure 4 by the solid vertical line.

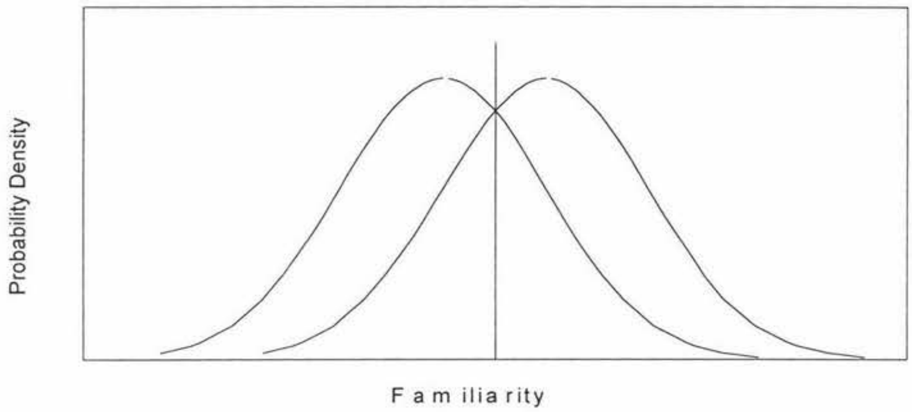

Figure 4: Representation of normal distributions of the familiarity of old and new words, with an unbiased observer's criterion located at their intersection.

Within word recognition experiments a variety of labels have been used to specify the psychological experience represented on x-axis in Figure 4. Word familiarity has been labeled "apparent oldness" by Bernback (1967), memory trace strength by Wickelgren and Norman (1966), and familiarity by Kintsch (1967, cited in Titus, 1973).

Figure 4 displays two normal distributions. The distribution with the higher mean represents the familiarity of the old words. The distribution with the lower mean represents the familiarity of the new words.

An unbiased observer will place the criterion at the intersection of the two distributions. However, participants can be either more liberal or conservative in their responding, placing the criterion more to the left or right, respectively, of the depicted line represented in Figure 4.

The most common parametric measures consist of the likelihood ratio (β), the criterion (c), and the relative criterion (c′). There are also a number of nonparametric alternatives to be discussed below. For a full summary of the most common response bias measures, the reader is referred to Macmillan and Creelman, (1990, 1996).

The original measure that was used to characterise response bias was β which is based on the principle of the likelihood ratio $l(x)$. This is the odds, or the likelihood, that either noise or signal plus noise was presented. In a recognition memory task it is thus the "ratio of the likelihood of obtaining an observation equal to the criterion given an old item to the likelihood of obtaining the observation given a new item" (Snodgrass & Corwin, 1988, p.36). It can be stated mathematically as:

$$\beta = l(x) = P(Y|y)/P(Y|n) \tag{5}$$

An unbiased observer will place the criterion at the intersection of the two underlying distribution, giving β=1. In a recognition memory task involving the presentation of previously seen words, β=1 will lead to the response strategy:

if $l(x) < 1$, respond new; if $l(x) \geq 1$, respond old

Because the likelihood ratio is based upon odds, it is assumed that the participants have certain pieces of information from which they can establish these odds. For example, participants need to know the a priori probabilities of the 'old' and 'new' words (Singh & Churchill, 1986). As a general rule, if $\beta=1$ this will result in optimal responding (McNicol, 1972).

Before analysis can be undertaken using β, it must first be transformed to produce interval scale data. To do this the natural logarithm is taken (Snodgrass & Corwin, 1988). After this transformation a neutral criterion is represented by a '0' score, liberal responding will produce a negative score ($\beta < 1$), and conservative responding will produce a positive score ($\beta > 1$) (McNicol, 1972).

However, the limitations of β have been addressed by a number of authors (see for example, Banks, 1970; Lockhart & Murdock, 1970; Macmillan & Creelman, 1990; Snodgrass & Corwin, 1988); therefore, the alternative parametric measure, c, is suggested.

While β establishes its position via the ratio of the heights of the new and the old distributions, c is established by means of measuring the distance between the intersection of the old and new distributions and the position of the criterion. c is defined as:

$$c = -0.5[z(H) + z(F)] \tag{6}$$

where *c* stands for the criterion. c is the average distance between the *z*-score for the hits and the *z-score* for the false alarms. The range of the criterion is the same as for d′ and is centered on 0.

Finally, there is the relative criterion, c′. This is a parametric measure where the criterion location is scaled relative to performance. That is, c is normalised by d′, so that it represents a proportion of the sensitivity distance. This is stated in the equation below (Macmillan & Creelman, 1990):

$$c' = c/d' \tag{7}$$

Macmillan & Creelman (1990) summarised several characteristics that are desirable in a measure of bias. The first three characteristics relate to isobias curves representing the position of all the points in ROC space that represent equivalent bias. An isobias curve is a predictor that indicates what will happen to performance when sensitivity changes and bias is constant (Macmillan & Creelman, 1991). Firstly, the bias measure should be monotonic with both the hit and false alarm rates. Secondly, the isobias curve should be able to measure bias sensibly both at the level of chance and below chance. The range of the measure of response bias should be independent of bias, and it should not be distorted when the scores are averaged across observers. It should be useful in theory and statistical analysis, and finally it should be associated with an index of sensitivity which is also desirable (Macmillan & Creelman, 1990).

With respect to these desirable qualities, several studies have shown that of the three measures, β, c, and c′, c is the most desirable, followed by c′ and then β. Therefore this study will adopt the response bias index, c. This measure is consistent with analyses undertaken by Hannay (1986), Macmillan and Creelman (1990), and Snodgrass and Corwin, (1988).

In addition to measures requiring parametric assumptions, so called "nonparametric measures" have been proposed. Hodos (1970, cited in Macmillan & Creelman, 1996) developed a triangle-based measure of response bias based on a similar geometric approach to that taken by Pollack and Norman for the development of their sensitivity measure, A′, already discussed. Hodos claimed this measure, termed B″, was assumption free. Using the triangular areas in ROC space depicted in Figure 3, Hodos defined B″ as follows:

$$B'' = (A_1 - A_2) / A_1 + A_2 + 2S \tag{8}$$

B″ is a function of the triangles S+A1 and S+A2 in Figure 3. The area of each of these triangles varies with the position of the obtained ROC point. According to this index, the closer the score is to -1, the more liberal the responding and the closer the scored is to +1, the more conservative the responding. Unbiased responding is reflected in a score of 0. This bias index was initially paired with the sensitivity index, A′.

In 1992, Donaldson suggested an improved measure of response bias, referred to as B''_D, which uses only the smaller triangles (A1 and A2) in Figure 3, rather than the larger triangles chosen by Hodos (1970). Thus, the equation:

$$B''_D = A_1 - A_2 / A_1 + A_2 \tag{9}$$

While these measures were identified as nonparametric because no parametric assumptions were used in their creation, this conclusion is not strictly accurate. Macmillan and Creelman (1996) showed that although there were no assumptions made in deriving the measures, they are consistent with underlying logistic distributions. Nevertheless, these so-called nonparametric measure will be used here to examine the data, so that the analysis is comparable to other published studies that have estimated bias from single point data using this approach.

Imagery

In this study, the imagery levels of the sets of words were altered in an attempt to vary d'.

Imagery is defined as a "mental representation of things that are not physically present" (Matlin, 1994, p.173), and it is generally assumed that visual imagery uses similar mechanisms to visual perception (Peterson & Graham, 1974).

Initially researchers thought that the meaningfulness or semanticality of words affected how well different words were remembered (thus affecting sensitivity levels). This assumption was based on the argument that, if words are more meaningful they will be attended to, and attention leads to encoding. Therefore words will be better remembered if they are meaningful. It was thought that content (open-class) words were more meaningful and thus more memorable than function (closed-class) words. However, researchers soon noticed that imagability (that is, the extent to which a word's referent could be examined in imagination), appeared to be a strong confounding variable. While the distinction between content and function words was still relevant, the most powerful variable turned out to be the imagery values of the words (Davelaar & Besner, 1988). For this reason, specification of imagery values became necessary.

Most of the available research into word imagery strength arose from the work of Paivio and various associates (Paivio, Yuille, & Madigan, 1968). Paivio (1986) instructed participants to rate words on a five-point imagery scale ranging from 'difficult' to 'very easy'. However, subsequently, Paivio et al. (1968) documented the use of a modified seven-point scale of 'low' imagery to 'high' imagery to eliminate narrow responding. The best way to obtain an indication of what are considered 'low' or 'high' imagery words, is to consider the instructions that were given to Paivio's participants:

"Nouns differ in their capacity to arouse mental images of things or events. Some words arouse a sensory experience, such as a mental picture or sound, very quickly or easily, whereas others may do so only with difficulty (i.e., after a long delay) or not at all. The purpose of this experiment is to rate a list of words as to the ease or difficulty with which they arouse mental images. Any word, which in your estimation arouses a mental image (i.e., a mental picture, or sound, or other sensory experience) very quickly and easily should be given a *high imagery* rating; any word that arouses a mental image with difficulty or not at all should be given a *low imagery* rating." (Paivio et al., 1968, p.4).

Thus, the imagery value of words is based upon the mental image that is produced by that word or set of words. For example, the word 'house' could create a fairly basic image of a house with a roof and chimney in one's 'minds-eye' without the object actually being present.

In total, 1,000 words were rated, and thus Paivio collected normative data for 925 nouns (the mean for imagery was 4.97 with a standard deviation of 1.93) with good validity. Support came from other work such as Devane (1988) who found that subjects were more likely to remember high imagery words than low imagery words in an unanticipated test. Further support came from the research of Ferlazzo, Conte, and Gentilomo (1993) who were interested in the effect of imagery values on event-related potentials (ERP). Their data confirmed that high imagery value words had a higher recognition rate than the low imagery words.

In view of the research just discussed it is expected that, in the current study, words with a low imagery rating will not be as easily recognised from a previously seen list as words that have a high imagery value. It is further assumed that the mean familiarity of the old words (x_o) will be higher than the mean familiarity of the new words (x_n), due to the participants' recent exposure to these words (Singh & Churchill, 1988). With reference to the three different imagery levels, it is expected that the low imagery words will produce means for the underlying familiarity distributions which are closer together than those for the medium imagery words, and the high imagery words will have the greatest distance between the means as reflected in the sensitivity index, (d').

Therefore, it is predicted that there will be an increase in the participant's d′ with the change in word imagery value (Hilgenhorf & Irving, 1978; Snodgrass & Vanderwart, 1980).

Dual coding theory

Paivio's theory has been included as it helps to explain the effects of imagery on memory. The dual-coding hypothesis proposed by Paivio is one theory that explains memory performance. In simple terms, the dual-coding hypothesis proposes that the human memory is comprised of two systems. One of them is based on verbal coding, and the other is based on visual imagery. It is assumed by this theory that the two processes are partially interconnected for tasks such as retrieval, storage and encoding, but essentially they are considered to be independent of each other. Because the two systems are interconnected in this manner, it stands to reason that one code can be transformed into the other. For example, when participants are presented with a word, this can evoke a non-verbal image, and when presented with a picture it can be named.

Since these two units are independent, this allows each system to be activated without activating the other system. It also means that they can both be simultaneously activated but at different levels, depending on the particular experimental task and the stimulus attributes. This enables the two codes to have an additive effect on participants' performance (Paivio, 1976; Paivio, 1986).

The theory states that there are three levels of processing, representational, referential and associative. Processing at the representational level occurs when linguistic stimuli form verbal representations and non-verbal stimuli form imaginal representations. Associative processing occurs when verbal stimuli activate verbal reactions and non-verbal images elicit other images. Finally, referential processing occurs when the established interconnection between the images and verbal representations are activated. This is displayed when an individual names an object that has been presented visually, or when a word stimulates a non-verbal image.

Referential processing is what occurs when a word forms an imaginal code. Words with a high imagery level arouse a non-verbal image with ease, while a low imagery words do so with difficulty. This process is assumed to be non-specific, as words can generate different images to different individuals, contingent on past experience and the current context (Paivio, 1976). For example, the word 'absolute' would not create a very clear non-verbal image to most individuals (as indicated by the norms, which give it a very low imagery value). On the other hand, a person who has had past experience with 'Absolute Vodka' may associate the word 'absolute' with their drink, thus eliciting a vivid non-verbal image and creating an artificially high imagery value.

Boles (1989) modified the original formulation of the theory as it could not explain all the results appearing in the literature. Boles noted that the problem was based around whether intentional or incidental learning had occurred in the experiment. From his research, Boles proposed that under incidental-learning word recognition tasks, a discrete model should apply. However, when intentional-learning word recognition is required, a continuous processing model should apply. Essentially then, these modifications of Paivio's theory (1971) made it a bimodal system which can adopt discrete processing when incidental learning occurs, and continuous processing when intentional learning occurs (Boles, 1989).

Imagery is also subject to individual difference. Firstly, some words may have specific attached meaning to some individuals, while for others no such association exists, as in the 'absolute' example above. Secondly, imagery can be actively employed by individuals and some may have more detailed imagery than others (Matlin, 1994). That is, individual imagery vividness may vary (Marks, 1973; Sheehan, 1967). For example, when told to imagine a house, 'low imagers' may develop a simple outline of a house, just covering the basics, (i.e., windows, doors, roof). In contrast to this 'high imagers' may have extreme detail on their house (i.e., the colour of the house, surrounding garden, details of the interior). Rehm (1973) noted the importance of visualisation in, for example, systematic desensitisation, where imagery is more beneficial, the more realistic the image is. Rehm was one of the first people to attempt to find an objective measure of visualisation variability.

Further support for the individual differences in imagibility skill and memory was found in a New Zealand study which investigated the differences in recall between high imagers and low imagers. Using the Vividness of Visual Imagery Questionnaire (VVIQ) participants were categorised into two groups, 'poor' and 'good' visualisers. A highly significant difference between the groups supported the hypothesis that individual image vividness is a good predictor of recall (Marks, 1973)[2].

Amnesics and Performance

Amnesia can result from numerous causes, many of which are related to age. Due to the overall increase in the average age of the population and thus in the predominance of amnesia, recognition memory has become of interest to researchers. In 1992 senile dementia was estimated to have affected over two million people in the United States of America (Bartus et al., 1982). In 1993, Alzheimer's disease was estimated to effect 5% of people of 65 years of age and 11% of those over 85 years of age. Other diseases associated with amnesia include Huntington's disease, which is rare and has a strong genetic basis, Parkinson's disease and, Korsokoff's syndrome. Aside from these causes, amnesia may also result from open or closed head injury (amnesia resulting from a closed head injury is usually referred to as post traumatic amnesia), or from medical lesions.

To understand impaired recognition memory, researchers must first discover how 'normal' memory works. Comparisons can then be made at a cognitive, neuroanatomical, and neurochemical level. From these comparisons researchers can attempt to define the problem and an attempt at some remediation can be made (Eichenbaum, 1994).

[2] Marks' results also found further support for the superiority of females over males on recall tasks.

Cognitive scientists have developed a multiple memory system model which neuroscientists are currently attempting to map onto different structures of the brain (Eichenbaum, 1994). A more reductionist approach addressed neurotransmitters such as acetylcholine (Snodgrass & Corwin, 1988). For example, research on Alzheimer's patients led to the development of the cholinergic hypothesis which states that lack of cholinergic markers causes dysfunctions in working memory (Bartus et al., 1982).

Typically a person with amnesia suffers from severe impairment of learning and memory for ongoing events (Warrington & Weiskrantz, 1968, 1970, 1974). According to the stage theory of memory (Atkinson & Shiffrin 1968, cited in Schacter & Tulving, 1994) for something to be remembered, it must first must first be encoded, then stored, and finally retrieved. Amnesia can be the result of a problem at any one of these stages, or all three of them (Eichenbaum, 1994). Two broad categories of amnesia are retrograde and anterograde amnesia. Anterograde amnesia is most typically seen, and is the inability to form new explicit memories.

Cohen and Squire (1980) concluded that in most amnesic patients' implicit memory – memory responsible for procedural tasks or skills – is spared, while explicit memory – declarative memory, which includes recall and recognition – is impaired. These events occur because structures such as the basal ganglia important for implicit procedural memories, are usually intact (Moscovitch, 1994), while the hippocampus, which is responsible for explicit declarative memory, is usually damaged (Moscovitch, 1994; Reed & Squire, 1997; Squire, 1994).

The actual mechanics of how or what exactly the hippocampus does is one of the most speculated topics in memory research (Johnson & Chalfonte, 1994; Shapiro & Olton, 1994). Theories about its function are, in reality, about the cause of amnesia.

According to some researchers, amnesics do not encode material as well as normal individuals, and thus base their judgments on an overall familiarity (Snodgrass & Corwin, 1988). These theorists propose that amnesic patients fail to encode the detailed specifics of a stimulus, and only encode the overall or global structure (Reintz, Verfaellie, & Milberg, 1996). Therefore, amnesics only remember certain stimulus parts, but not how those parts are interrelated. Remembering these parts in their exact structure is referred to as the process of 'binding'.

Support for the theory that amnesics suffer a breakdown in the process of binding has come from experiments such as those carried out by Reintz (1996) and his associates who were investigating memory conjunction errors using words (for example; handstand, shotgun, handgun). Overall, it was concluded amnesics failed to discriminate as well as the control subjects, mostly on conjunction stimuli, which produced the highest false alarm rate. This result lends support to the hypothesis of a lack of specific encoding because the amnesics were aware of the stimulus parts, but not their specific combination. This theory implies that amnesia is due to problems with the encoding of the information (Baddeley, 1994) and this is the view that Snodgrass and Corwin (1988) adopted to explain amnesia.

Chapter 2
Measures and Expectations

The study

In the current experiment a standard recognition memory task was used. The methodology was based upon Experiment 1 of Snodgrass and Corwin (1988). The main differences and similarities between the two studies are outlined below.

Both studies allocated approximately the same amount of time (two minutes exposure) to the presentation of the memory list. In each case this list comprised 60 words. However, in Snodgrass and Corwin's (1988) experiment, half of the words were high-imagery and half were low imagery words. In the present study, all 60 words were in only one of the imagery categories. Both studies allowed the participant to set their own pace when deciding if the words are 'old' or 'new' and both used a method of random selection for the order of presentation of the words.

There were several differences between the studies. Firstly, the mode of presentation differed. In Snodgrass and Corwin (1988) the words were presented via index cards (measuring 3x5 inches) upon which the words were typed in uppercase letters. By contrast, in this study the mode of presentation was via a computer monitor. Secondly, the experiments differed in the number of trials per d′ estimate. In this study, three blocks of 120 trials were undertaken at each imagery level, so that individual bias and sensitivity estimates were based on 360 trials, whereas Snodgrass and Corwin used 30 trials per estimate. Thirdly, in this experiment testing occurred at three imagery levels; low, medium, and high, while Snodgrass and Corwin used two: low and high. Finally, Snodgrass and Corwin had a filled two-minute delay between the presentation of the memory list and the recognition task, during which the payoff matrix was explained. In the present study an 18-second intermission occurred during which instructions on how to respond in the testing session were reiterated on the computer screen.

Bias in Recognition memory tasks

Snodgrass and Corwin (1988) noted that normal participants were, on average, more conservative, while amnesics were more liberal. Their conclusions have since been supported by the work of Reintz et al. (1996). They also concluded that amnesics were more likely to respond that a completely new stimulus was old in a conjunction errors experiment[3]. Further still, work by Gardiner and Gregg (1997) who were researching the remember-know paradigm, noted that the elderly participants who were liberal in their responding were actually more confident in their recognition responses than the younger participants, even though this confidence was misplaced. Overall, there is a general consensus that this is due to the high level of familiarity that is evoked by the stimuli.

[3] A conjunction-errors experiment uses four categories of nonsense words defined as follows: Old words are nonsense words comprising a memory list, new words are previously unpresented nonsense words, conjuction words are constructed from parts of the old words, and feature words are constructed from parts of both the old and the new words.

Other research comparing 'amnesics' and 'normals' has indicated that liberal bias in amnesics is quite consistent (Butters et al., 1985) and others have even claimed that false alarm rates are almost as effective as d′ for discriminating between Alzheimer's and normal elderly populations (Branconnier, Cole, Spera, & De Vitt, 1982, cited in Snodgrass & Corwin, 1988). It is from this observation that Snodgrass and Corwin (1988) concluded that different types of bias could represent different types of memory. For example, demented patients are the most liberal in their responding, followed by amnesics and lastly normal participants. However, these conclusions can only be made if both measures of sensitivity and bias are independent, and if an individual's responding behaviour is independent of sensitivity. If response behaviour is affected by sensitivity, this could have implications for conclusions drawn on the basis of other research. Snodgrass and Corwin acknowledge that the same level of sensitivity was not achieved for both groups. Therefore, the more liberal responding could be due to different sensitivity levels as opposed to levels of amnesia.

Research Aims

It is not certain how systematic changes in sensitivity effect response bias. For this reason, a comparison between the response biases of groups or individuals who differ markedly in sensitivity may not be valid. Therefore, the aims of this research are:

1) To investigate whether a relationship exists between sensitivity and bias. Specifically, bias will be examined as imagery level is manipulated so as to alter sensitivity.

2) To investigate whether the relationship between bias and sensitivity is consistent across individuals.

3) To compare the influence of sensitivity on naturally occurring and externally imposed biases. Externally imposed biases will be manipulated through changes in a payoff matrix specifying the consequences of correct and incorrect responses.

Chapter 3
Experiment 1

In this experiment participants' natural biases were explored. The experimental design ensured that there were no extrinsic reasons to adopt a response bias, because old and new words were presented with equal probability.

Method

Participants

Eight volunteers participated in the study. The group comprised four men and four women, all European and aged between 22 and 57 years of age. One participant had English as a second language but was fluent in English. Four were students or staff at Massey University Albany, and two were students at Auckland University. The remaining two were in full-time employment outside of the university. Two further individuals volunteered but did not complete all the blocks of trials. Their data were not included in subsequent analyses.

Materials and apparatus

Word list

The words to be used in the research needed to have a validated imagery rating. To obtain the necessary number of words categorised according to imagery level, the MRC (Medical Research Council) psycholinguistic database was consulted.

The MRC psycholinguistic database was established by Max Coltheart in order to give researchers the information they needed to minimise the effect of confounding variables that arose because of the particular properties of words, such as irregular spelling, imageablilty/concretness, and frequency in the English language (in written form). The database is comprised of three principle files, the DICT file, the R-S file (responses from word-association experiments, including information about the words used and the words evoked as a response), and the S-R file (details on word-association responses to a set of words).

Of particular interest for this research is the DICT file. This is a dictionary which includes information about the semantic, orthographic, phonological, and syntactical properties of a set of words. In 1981 Coltheart reported that the file contained a massive 98,538 words, but not all of them had a rating on all 27 different linguistic properties that the file encompassed. For example, imagery ratings had been obtained for only a subset of 8903 words (Coltheart, 1981). In 1998 when the MRC database was consulted for this research through an Internet web site, the number of words had increased, and one of the psycholingusitic properties had been dropped. At the time the database was accessed the total number of words was 150,837 rated on 26 linguistic properties. Of these words, 9,240 had an imagery rating.

The information for this database was obtained by compiling eight smaller databases (the SOED - Shorter Oxford English Dictionary, the Edinburgh Associative Thesaurus, the Colorado Norms, the Gilhooly-Logie norms, the Paivio Norms, the Kucera-Francis frequency count, the Thorndike-Lorge frequency count, and finally, Daniel Jones' phonetic transcriptions of the Pronouncing Dictionary of the English Language, 12th edition). The imagability scores were obtained by merging of the Gilhooly-Logie, Paivio, and Colerado norms. The range of imagery ratings for this set of words is 129-669, with a mean of 450 and a standard deviation of 108.

Words for three imagery levels, low, medium, and high were selected and then semi-randomised into the nine versions of the program (3 x imagery levels x 3 different payoff matrices). A single disk contained three variations of the program. The difference between the programs was in the content of their word lists, either low, medium or high imagery. The set of low imagery words consisted of a total of 1080 words (360 per program) and had a range of imagery values between 143-376. The range of medium imagery values for the set was from 417 to 501. Thus, there was a gap of only 40 imagery values between the top of the low and the bottom of the medium ranges. The range of imagery values for the high set was from 548 to 667. The gap between the top of the medium and the bottom of the high imagery ranges was 47 units. Due to a shortage of words available, especially in the low imagery range, and the desirability of having no overlap between the imagery values for the three conditions, it was decided that words up to (and including) ten letters long could be used.

Testing venues

For the convenience of volunteers recruited from outside of the university, testing was carried out at two locations, at each of which purpose written software (using Turbo Pascal v6.0) was run on a P.C. Each subject was tested at only one location.

Environment One

Environment one was a room which measured 2060 mm by 2740 mm. When testing was in progress wooden blinds were closed and artificial lighting was used. This consisted of two 'spot lights', both containing 100 Watt bulbs, and one 60 Watt bulb in the middle of the room, which was covered by a frosted shade. In this room testing was conducted using a PC-Direct 500 hard drive, and a PC-Direct monitor with a screen measuring 275 mm by 200 mm. Participants sat approximately 680 mm from the screen. Five participants completed the experiment in this environment.

Environment Two

Environment two was smaller than environment one, with the room measuring 1890 mm by 1660 mm. In this room, testing was conducted using a 486SX 25 Advantage Computer with a VGA Phillips monitor with a screen measuring 275mm by 200mm. Participants sat approximately 640 mm from the screen, with the chair positioned to coincide with markers on the floor. Three participants completed the experiment in this environment.

Environment two contained only one light fitting which was a fluorescent tube in the middle of the room. There was very little natural light in this room.

Procedure

Participants were tested individually in one of the testing venues. On the first session the experimenter explained the procedure, using written instructions contained in individual folders. The folders contained details of the experiment, and a pre-prepared order through which the participant proceeded during subsequent testing sessions. The first block of trials was completed while the experimenter waited in a room nearby.

A maximum of two blocks of trials could be completed per day, with a minimum of two hours between the two blocks. Otherwise trial blocks were completed at the participant's convenience. Participants received 'incentives' (such as chocolate and sweets) on a variable ratio schedule to help maintain motivation.

Participants accessed the appropriate computer program and selected the correct file by consulting their 'grid', which comprised a set of instructions discussed with them at their first session (included in Appendix A).

In the first half of the test a total of 60 words (the memory list) were presented in a random order. It took two minutes to present the memory list (approximately two seconds per word). The words were placed in the center of the screen in a white font on a black background. Upper case letters were used and subtended a visual angle of approximately 0.42° in Environment 1 and 0.45° in Environment 2.

At the completion of presentation, a new screen appeared with instructions on how to respond in the recognition test. The 'B' and 'N' keys were used as the response mechanism, with the 'B' key (left hand) used to signal 'yes, seen previously' and 'N' key (right hand) used to signal 'no, not seen previously'. This screen stayed on for 18 seconds before the recognition test began.

During the test, a total of 120 words were presented, 60 of which had been presented as the memory list, plus 60 distractor items. For each participant, on the first block of trials at each imagery level, the computer made a unique random selection to determine which words would be 'old' or 'new'. This selection determined the 'old' and 'new' words for all three blocks of trials at that imagery level. With one inadvertent in Experiment 2 mentioned later, each word was used only once with each participant. The testing was self paced, in that next trial was initiated only after a response was made at the keyboard.

At the completion of each trial, the programme displayed a summary screen which included details of the number, and the proportion, of hits and false alarms. Data generated from the trials were automatically written to disk under the participant's ID.

Results

Estimates of bias and sensitivity were calculated separately for each participant at each imagery level. Software was written to estimate both parametric and so-called nonparametric indices. The program took hit rates and false alarm rates as input and estimated d′, c, the area under the ROC, and Donaldson's bias measure, B''_D. The z-scores required for the calculations were based on Odeh and Evans' (1965) algebraic approximations (out of Macmillan & Creelman, 1991). The estimate of the area under the ROC, A′, was obtained using Pollack and Norman's (1964) method. Although entire ROC curves were not obtained, parametric measures were used (as mentioned previously) as they are most strongly supported in recent literature, and were used in Snodgrass and Corwins' (1988) research. So-called 'nonparametric' measures were also used because many researchers consider these preferable when the estimates are based on only a single point on the ROC. These measures are thus reported here for comparison with other published studies.

Sensitivity

The d′ and A′ values at each imagery level, averaged across participants, are shown in Table 1, along with their standard deviations. As expected, as imagery level increased, there was a tendency for sensitivity to increase. Individual data are included in Appendix C.

Table 1. Means and standard deviations of individual d′ and A′ across imagery levels.

	Imagery Level		
Measures	Low	Medium	High
d′	1.58 (.42)	1.90 (.53)	2.25 (.58)
A′	.85 (.05)	.88 (.05)	.91 (.05)

Separate repeated measures analyses of variance (ANOVA) were conducted on the individual d′ and A′ measures using Statistical Analysis System (SAS, 1994, Carry, NC) software. The three levels of imagery comprised the independent variable in each case. At an alpha level of $p<.05$ the main effect of imagery was significant for both parametric and nonparametric measures. Details are as follows: for d′, $F(2,14) = 23.36$, $p=.0001$, for A′, $F(2,14) = 12.07$, $p=.0009$. Post hoc tests using Tukey's HSD, indicated that the significant difference in d′ was between the low imagery and high imagery words. d′ for medium imagery words was not significantly different from either high or low imagery words. On the other hand, the comparable post hoc tests with the nonparametric measures, confirmed a significant difference between A′ at all three imagery levels.

Response Bias

Similar repeated measures analyses to those used with the sensitivity measures were conducted on the two indices of bias, c and B''_D. Imagery remained the independent variable and c and B''_D were the dependent variables. The means and standard deviations of these measures are summarised in Table 2.

Table 2. Means and standard deviations of c and B''_D across imagery levels.

Measures	Imagery Level		
	Low	Medium	High
c	0.32 (.27)	0.41 (.39)	0.38 (.36)
B''_D	.23 (.37)	.46 (.43)	.43 (.63)

There were no statistically significant results from the ANOVA at an alpha level of $p<.05$, the details of these results being as follows: for c, $F(2,14)=.41$, $p=.67$, for B''_D, $F(2,14)=.84$, $p=.45$. The lowest means and standard deviations were for low-imagery words.

A tendency, rather then a firm relationship between bias and sensitivity was expected, and so the relationship between sensitivity and bias for the data as a whole was considered further. The group data are graphed in Figures 6 and 7 below. Each graph displays bias as a function of sensitivity. The dashed lines in the figure represent unbiased responding. The data points represent individual data (individuals are not specified) at each imagery level. The different imagery levels are represented by the different symbols on the graph. The abscissa on Figure 5 representing d' has a range of 0-6. This range was used to enable later comparisons with Figures 8 and 10 from Experiment 2. From Figure 5 it can be seen that criteria are mainly positive, although the range is between -.141 and .836. The graph appears to suggest a negative relationship between d′ and c which presents itself most strongly for high imagery, and slightly for medium imagery, but is not evident for the low imagery words.

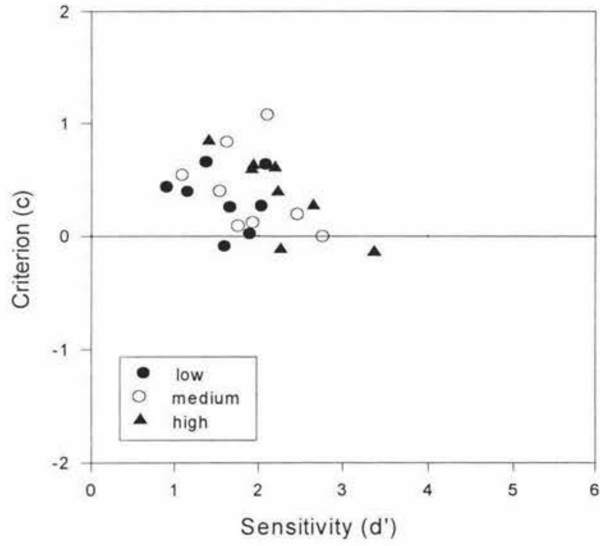

Figure 5. The criterion c, as a function of d′ at all three imagery levels.

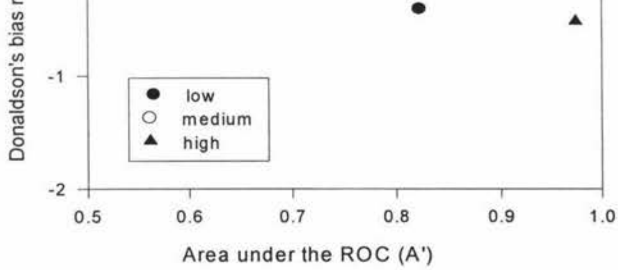

Figure 6. The criterion B''_D, as a function of A' at all three imagery levels.

Figure 6 illustrates the relationship between the nonparametric indices of sensitivity (A′) and bias (B''_D). The x-axis has been drawn to extend over the entire range of possible values. This figure confirms the pattern evident in Figure 5. The suggestion of a negative correlation between sensitivity and bias becomes more obvious in this graph, and again, it is more prominent for medium and high imagery, and less obvious for low imagery. To investigate the implied negative corrlations formally, Pearson product-moment correlation coefficients were calculated between bias and sensitivity. A summary of these results is presented in Table 3.

Table 3. Correlation coefficients between measures of response bias and sensitivity under three imagery levels.

	Imagery Level		
Correlated Variables	Low	Medium	High
d′ and c	-.18	-.36	-.82*
A′ and B''_D	-.07	-.74*	- .92**

** $p< .001$.
* $p< .05$.

As can be seen in Table 3 all of the correlations are negative, and there is a significant strong correlation between bias (c or B''_D) and sensitivity (d′ or A′) for the high imagery words, and a significant moderate correlation between B''_D and A′ for the medium imagery words. A correlational analysis of the relationship between sensitivity and bias regardless of imagery level, produced a significant result ($p<.05$) for the relationship between A′ and B''_D (r = -.46, p=.023) but not for c and d′ (r = -.40, p=.053).

Discussion

The attempt to provide variation in performance by varying imagery levels of words in a word recognition task was moderately successful, as the ANOVAs indicated that both the parametric and nonparametric estimates of sensitivity were significantly different as a function of imagery level, except that the medium imagery data were not distinguished from the other levels when parametric measures were used.

Informal inspection of the individual data revealed that for seven of the participants, d′ was increasing across imagery levels, while for the remaining participant d′ peaked at medium imagery. However, only two of the seven showed a marked increase in d′ across imagery levels, while the remaining five varied little. This lack of a dramatic difference across the imagery levels is likely to underlie the ANOVA result for this parametric index. Nevertheless, perusal of the means in Table 1 indicate that for both sensitivity measures, (d′ and A′) the trend was for mean sensitivity to increase as the imagery level increased.

The intention in Experiment 1 was to explore changes in response bias with changes in sensitivity. However, the changes induced in sensitivity were relatively weak and would therefore be unlikely to reveal such a relationship unless it was a particularly strong one. The correlational analysis, however, provided some support for the existence of a relationship. At the low and medium imagery levels, no significant correlation was obtained between c and d′. At the high imagery level, however, a strong negative relationship was obtained. On the other hand, for the relationship between A′ and B''_D, significant correlations were obtained at the medium and high imagery levels. The high imagery level produced a very strong negative relationship.

The data suggest that response bias became less pronounced as sensitivity increased, which is in keeping with the findings of Stillman, Brown, and Troscianko (unpublished manuscript, 1998) in sensory studies in taste and hearing.

Summary

A clear finding of Experiment 1, which supports the findings of Snodgrass and Corwin (1988) was that where 'old' and 'new' words were presented with equal probability, so that there was no pressure to adopt a response bias, participants generally displayed a tendency towards conservative responding. This is best illustrated in Figures 5 and 6, where responding relative to the point of no bias is illustrated. Data points above that zero bias line ($y > 0$) represent a conservative bias, while points below represent a liberal bias ($y < 0$). Therefore it can be seen on these graphs that responding by the group was quite conservative.

The outcomes of Experiment 1 suggest a tendency for participants to be more conservative in their responding as sensitivity decreases. This result is interesting in light of the results of Snodgrass and Corwin (1988). In their study the level of difficulty was not matched between the three groups (amnesic, demented and normal) rather the normal group found the task easier than the remaining two groups. However, the amnesic and demented groups responded in a robustly liberal manner, and became increasing liberal the more trials they completed. The results of the current study would predict that, as the task was harder for these groups, their responding should have been more conservative. In a speculative vein, it may be that a group's natural bias becomes more extreme as the difficulty of the task increases. Therefore, normal participants would become increasingly conservative and amnesic or demented participants would become increasingly liberal in a parallel fashion with task difficulty. To find support for this theory, this study or one similar to it, would need to be repeated using either amnesic or demented volunteers as the participants.

Chapter 4
Experiment 2

The aim of the Experiment 2 was to explore the relationship between bias and sensitivity when participants were induced to adopt either a more liberal or conservative response bias than they were naturally predisposed to. To attempt to alter response bias, the four different response outcomes shown in the detection theory response matrix (Figure 1) were allocated different costs or rewards, thus creating a weighted payoff matrix (Macmillan & Creelman, 1991). Various reinforcers such as money or points can be used as rewards. In this study, points were allocated to each response in the same manner as Snodgrass and Corwin (1988), who successfully altered response bias.

Manipulating the payoff matrix for different response outcomes allows the relationship between sensitivity and bias to be investigated further. Experiment 1 was aimed at assessing this relationship under a natural bias, while Experiment 2 explored whether the same relationships occurred when bias was imposed externally.

An additional aim was to provide data which, in combination with that from the same individuals in Experiment 1, could be used to obtain maximum likelihood estimates of the parameters of the ROC. These would enable the single-point estimates of d′used in Experiment 1 to be replaced with more accurate estimates against which to assess response bias.

Method

Participants

The volunteers who participated in Experiment 2 were the same as those in Experiment 1.

Materials

The materials used in this experiment (environments, program base, and word list source) were largely the same as those used in Experiment 1. Variations to the materials used in the experiment included additions to the folder (see Appendix B), consisting of instructions and information sheets as to the aim of Experiment 2, details of the payoff matrices, and a scoring guide. New versions of the computer program used in Experiment 1 were created using the same ranges of imagery values for the high, medium, and low sets to create new lists of words for use with payoffs promoting both conservative and liberal response biases.

Procedure

The basic procedure for Experiment 2 was the same as for Experiment 1 however, the order of the liberal and conservative payoffs was counterbalanced, with participants completing all the trial blocks for one payoff before commencing the other. In line with

Experiment 1, prior to the commencement of each new payoff, participants met again with the experimenter to ensure that the new instructions were understood.

The responses made by the participants were associated with points which were awarded contingent on appropriate use of the payoff matrix. Under the liberal payoff, points were allocated as follows:

10 points were given for each hit,

1 point for each correct rejection,

-10 points for each miss, and

-1 point for each false alarm.

Whereas, under the conservative payoff, points were allocated in the opposite manner:

-10 points for each hit,

-1 point for each correct rejection,

10 points for each miss, and

1 point for each false alarm.

The points attained in each session were recorded by the computer and displayed on the screen at the conclusion of the trial block.

Two of the participants who completed Experiment 2 received two explanations for the conservative and liberal conditions (compare in Appendix B). Originally the instructions indicated that participants should answer in the same manner as in the 'none' condition, but when in doubt they should take into account the payoff matrix. However, the other form (Appendix B) made it clear that the aim was to get as many points as possible under the payoff matrix. Changes were made because one participant concluded that the best score would be obtained by responding either only 'old' or only 'new' depending on the payoff. For this reason testing was stopped for three days and the old form was replaced with a new form. The second participant began Experiment 2 using the new form The participant who had conducted a session by responding only 'new' in the conservative condition, was asked to take a break for a week, and he then restarted, beginning with the liberal condition. The remaining participants began the experiment using the revised instructions.

In addition to the incentives offered in Experiment 1 (which continued throughout Experiment 2), an overall prize was offered for the participant who accumulated the most points in the liberal and conservative conditions combined.

Results

Estimates of sensitivity and bias were obtained in the manner reported in Experiment 1. Both parametric and nonparametric measures were gathered. Analysis was conducted using SAS software.

Liberal Bias and sensitivity

The bias (parametric and nonparametric) measures are shown graphically as a function of sensitivity in Figures 7 and 8 respectively.

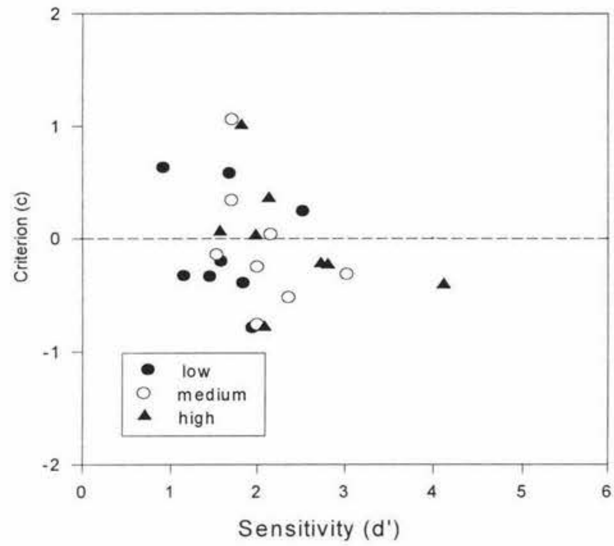

Figure 7. The criterion, c, as a function of d′ at all three imagery levels under the liberal payoff matrix.

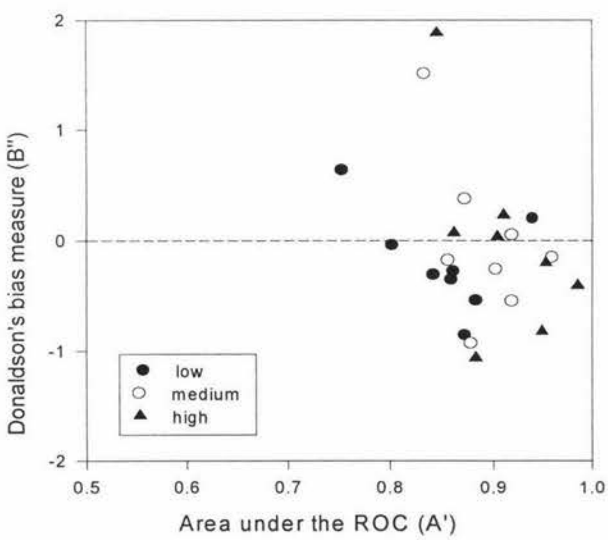

Figure 8. The criterion, B''_D, as a function of A′ at all three imagery levels under the liberal payoff matrix.

Figure 7 illustrates the parametric results when the external liberal bias was imposed. The criteria are quite diverse, ranging between 1.062 and -.788. Figure 8 displays nonparametric data. Separate repeated measures ANOVAs were conducted with imagery as the independent variable and c and B''_D as the dependent variables. No significant result ($p<.05$) was found for either c or B''_D.

As mentioned previously in Experiment 1, because a tendency as opposed to a firm relationship was expected, correlations again were calculated between the two variables. These results are summarised in Table 4.

Table 4. Correlation coefficients between measures of response bias and sensitivity at three imagery levels when liberal bias was imposed.

	Imagery Level		
	Low	Medium	High
Measures			
d′	-.18	-.43	-.45
A′	-.45	-.49	-.55

The obtained correlations between these variables were low, and none was significant at $p<.05$. A similar result was obtained when the correlations were determined irrespective of imagery level.

Conservative Bias and Sensitivity

Similar ANOVAs were conducted here as for the liberal data. Again, no significant results ($p<.05$) were found for either c or B''_D. The data are graphed in Figures 9 and 10 which represent the individual bias measures as a function of sensitivity. The conservative parametric data (as graphed in Figure 9) displays criteria that lie almost exclusively above the no bias line, with a range of 1.515 to -.018. Figure 10 illustrates the nonparametric measures obtained under the conservative payoff. Again the points mostly lie above the no bias line.

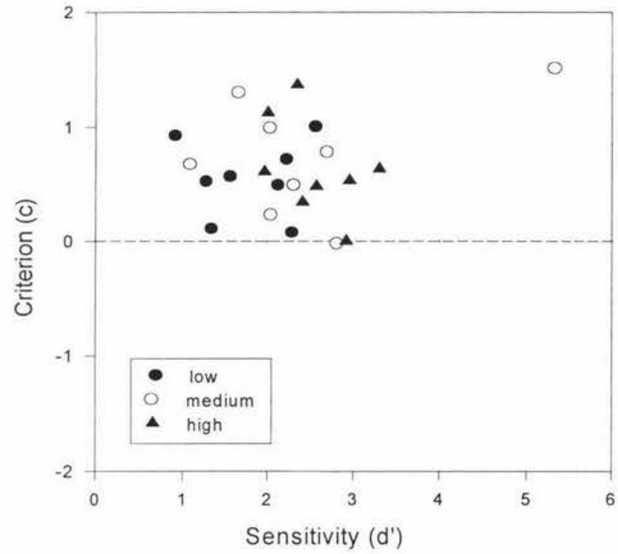

Figure 9. The criterion, c, as a function of d′ at all three imagery levels under the conservative payoff matrix.

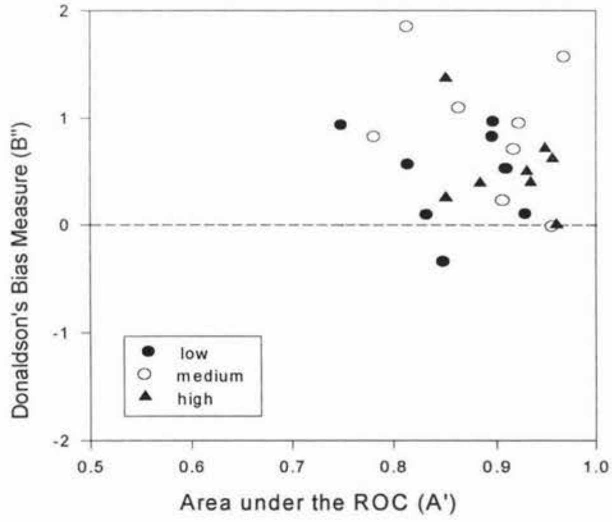

Figure 10. The criterion, B''_D, as a function of A' at all three imagery levels under the conservative payoff matrix.

As for the 'liberal' payoff matrix, correlations were calculated on the parametric and nonparametric results for the conservative data to further investigate any relationship between sensitivity and bias. These results are summarised in Table 5.

Table 5. Correlation coefficients between measure of response and sensitivity at three imagery levels for the conservative payoff matrix.

	Imagery Level		
Correlated Variables	Low	Medium	High
d' and c	.03	.38	-.45
A' and B''_D	-.16	-.32	-.76*

* $p < .05$.

Four of the six correlations were negative. The only significant result was a moderate correlation between B''_D and A' for the high imagery words. Correlations for the data as a whole, calculated regardless of imagery level were not significant. The details are as follows, between d' and c, $r = .19$, $p = .38$, A' and B''_D, $r = -.26$, $p = .29$.

Whether or not sensitivity varies with an externally imposed response bias requires the assumption that the 'imposed' bias has been adopted by the participant. Therefore, the liberal payoff should lead to a criterion which is lower than the natural bias, and a conservative payoff should lead to a criterion which is higher than the natural bias.

To asses if bias had been significantly altered, again repeated measures ANOVAs were conducted with payoff matrix as the independent variable, and c and B''_D as the dependent variables. Data are included here from Experiment 1 as it is necessary to check for a comparative shift from the participants' natural bias. In Table 6, which summarises the means, this is labeled 'none' (representing no externally imposed bias).

Table 6. Means of c and B''_D across the three payoff matrices for low, medium and high imagery words

Measure		Payoff		
		Liberal	None	Conservative
c	low imagery	-0.07	.32	.56
	medium imagery	-0.07	.41	.75
	high imagery	-0.03	.38	.64
B''_D	low imagery	-0.19	.23	.47
	medium imagery	-0.01	.46	.91
	high imagery	-0.03	.43	.82

At an alpha level of $p<.05$, the main effect of the payoff matrix on response bias was significant for all three imagery levels. For low imagery, the details were as follows: c, $F(2,14)=13.72$, $p=.005$; B''_D, $F(2,14)=11.93$, $p=.0009$. For medium imagery: c, $F(2,14)=9.17$, $p=.0029$; B''_D, $F(2,14)=8.72$, $p=.0035$, and for high imagery; c, $F(2,14)=13.79$, $p=.0005$; B''_D, $F(2,14)=14.56$, $p=.0004$.

The post hoc test (Tukey's HSD) indicated that for both c and B''_D the significant difference with low and high imagery words was between the bias adopted with a liberal payoff and that adopted with either no payoff or a conservative payoff, but there was no significant difference in bias between the already rather conservative natural bias and that adopted under the conservative payoff. For medium imagery, the post hoc test indicated that the significant difference was between the bias adopted under the conservative and liberal payoffs. Participants' natural bias did not differ significantly from either the 'liberal' or 'conservative' biases.

While there may be reasons why bias changes with sensitivity, there should not be any variation in sensitivity with bias. Any such variation should most probably be attributed to an unequal distribution of certain dimensions of the memory list (for example, familiarity and imagery). To check variation in d′ across similar lists, one final analysis was run. This was aimed at testing that d′ was equivalent across the three payoffs. An ANOVA was run where the 'payoff' (liberal, none, conservative) was the independent variable, and the d′s for the individuals at the three imagery levels was the dependent variable. The only significant difference between the groups was for imagery level. The main effect of payoff was not significant.

An additional aim for manipulating response bias was to provide data which, in combination with that from Experiment 1, could be used to obtain a maximum likelihood estimates of the parameters of the ROC. It was hoped that this would then allow the less accurate single-point estimates of d′ used in Experiment 1 to be replaced with more accurate estimates against which to assess bias.

Dorfman and Alf's (1969) algorithm was used to produce maximum likelihood estimates of the slope and the intercept of the ROC from the three points on the individual ROCs available at each imagery level as a consequence of varying response bias. Usually a rating procedure is used to obtain data, thus allowing a larger number of points to be obtained. However, a rating procedure was not considered to produce a suitable indication of response bias in this case. In addition, using the binary method ensured that the results would be comparable with those of Snodgrass and Corwin (1988).

Unfortunately the variation in bias with the three payoffs was inadequate for sensible maximum likelihood fits, as indicated by the mostly very shallow slopes for the fitted ROCs. Inspection of the data indicated that no improvement in the accuracy of d′ estimates was likely as a consequence of the maximum likelihood fits. Under the normal-normal-equal variance model a slope of 1.0 is expected for the ROC on z co-ordinates, reflecting the equality of variance.

An arbitrary decision was made to include only those fits for which the slopes were within $\pm$ 0.2 of 1.0. Under this criterion no fits were accepted (see Appendix C). Inspection of the individual data, however, suggested that in this case the slopes were a consequence of insufficient or inconsistent variation of the criterion with changes in payoff.

Discussion

The aim of Experiment 2 was to further investigate the empirical question of the relationship between sensitivity and bias by externally imposing a response bias. It was not unexpected that little relationship would be apparent as in Experiment 1 as participants were asked to adopt particular biases. Therefore, any natural tendency for a change in bias with a change in sensitivity is likely to be overridden as the participants attempted to comply with the instructions given to them.

While no significant results were found there were some interesting trends. Firstly, the results indicate that if there is a relationship between sensitivity and bias with respect to behaviour, it is a negative relationship. When liberal or conservative bias was imposed, correlations between sensitivity and bias (d′ and c, B''_D and A′) were mostly negative. None, except the high imagery nonparametric measure was significant.

Interestingly, imposing bias externally was not as successful as the results obtained by Snodgrass and Corwin (1988) even though the same payoff matrix was used in a similar task. However, overall it can be safely concluded from this analysis that response bias was significantly different under the liberal and conservative payoffs.

The conservative trends in responding evident in Experiment 1 are confirmed by the fact that the biases adopted naturally by participants were (statistically) the same as those adopted under a payoff favouring conservative responding. This is also reflected in Figures 5 and 9.

One disappointing outcome of this experiment was the lack of sensible data from which maximum likelihood estimates could be made. For this reason, it was not possible to replace the single-point estimates of d′ as originally planned.

Chapter 5
General Discussion

The aim of this research was to investigate the relationship between sensitivity and response bias, either natural or externally influenced. The study was prompted by the results of two pieces of work. Firstly, the work of Snodgrass and Corwin (1988) who proposed that different groups (amnesics, demented, and normal participants) could be distinguished by the type of response bias they adopted. Secondly, the more recent work of Stillman, Brown and Troscianko (unpublished manuscript 1998) obtained some evidence that individuals' response bias may be affected by sensitivity. If the findings of Stillman et al., in the modalities of taste and hearing apply to recognition memory as well, they may have implications for the conclusions of Snodgrass and Corwin. In the latter study, sensitivity was not matched across all three groups. Therefore, their conclusions concerning the relative characteristic biases adopted by the different groups might be somewhat premature. With this in mind, it was decided to investigate whether a relationship or trend could be observed between sensitivity and response bias in a recognition memory task similar to that used by Snodgrass and Corwin (1988). Assessment of this relationship was achieved by investigating how participants' natural bias 'behaved' as sensitivity level was altered for the same individuals via manipulation of word imagery levels.

Manipulation of sensitivity via imagery levels

In line with previous research findings (Devalaar & Besner, 1988; Devanne, 1988; Hilgenhorf & Irving, 1978; Paivio et al., 1968; Singh & Churchill, 1988; Snodgrass and Corwin, 1988), as imagery level increased, the recognition-memory task become easier, reflected in an increase in the measure of sensitivity.

Analyses confirmed that the area measure of sensitivity, A′, was statistically different between all three imagery levels. However, the sensitivity index, d′, was only significantly different between the low and high imagery levels. The lack of a clear difference in d′ between all three levels was disappointing. This could have resulted from the content of the word lists themselves. The shortage of words with a rated imagery level, and the desirability of having three distinct levels (low, medium, and high) resulted in the inclusion of some words which, in hindsight, would have been better omitted. Firstly, in the low imagery set, approximately 26% of the words were considered 'novel' (where novel is defined as having a rating of zero in the Brown Frequency[4]). The novelty of these words may have artificially raised individual d′s for the low imagery words because their uniqueness may make them more memorable. Examples of these words include 'huzza', 'tush' and 'nabob'. Some of the low frequency words may have appeared to have been typing mistakes, for example, 'bhang', and may have been memorable on that account.

[4] Brown Frequency scores are an indication of how often words occur in 'every day life'. The Brown Frequency scores represent how many times a particular word occurs in the London-Lund Corpus of English Conversation.

The effect of word frequency on recognition has been well documented ever since Brown (1976) first proposed that familiarity was not the only dimension upon which word recognition decisions were based. Rather, he considered that individuals also took into account how memorable words were. One of the factors that was found to contribute to memorability was word frequency (Brown et al., 1977).

Further research into word frequency led to the discovery of the "mirror effect" which is that certain properties of words increase the accuracy of recognition as they are more likely to be recognised as old when old, and new when new (Glanzer & Adams, 1990). Effectively, this is what occurs with words of low frequency (Maddox & Estes, 1997; Shiffrin & Steyvers, 1997), that is, words which have a low frequency score will be more easily judged as 'new' when presented on the distractor list, and will also be easily judged as 'old' if they have occurred in the memory list. While some researchers rated the influence of word imagery and word frequency equally (Christian, Buckley, Taraka, & Clayton, 1978), according to the work of Paivio (1971), imagery is more influential on word recognition. Therefore, it is highly probable that word frequency was a confounding variable, and it may have hindered the effect of imagery level. It is proposed that the larger proportion of low frequency words in the low imagery set artificially raised sensitivity levels, thus bringing the levels associated with the low and medium imagery word lists too close to be distinguished on a statistical analysis.

Aside from the influence of novel words, there were errors made in the entry of the words into the programme. Firstly, some of the words were entered under their American spelling, for example, the word 'centre' was entered as 'center' and this may have affected the familiarity of the word. Spelling errors or 'typos' were also made. Overall, from all nine programmes there was a total of 0.77% spelling errors or typos (25 out of 3240 words), and 0.12% (four) of the words were spelt the 'American Way'. Finally, 0.15% (five words) were repeated (entered twice) in the one programme. These errors are more than likely to have affected the memorability of those words. However, inspection of the nine individual programme lists indicated that the errors were roughly equally distributed across the three word imagery levels.

The outcome of the study suggests that the separation of the three word sets in term of imagery level was less than ideal. Ideally, there should have been a larger gap in imagery values between each. However, due to the shortage of image-rated words and the need to test across as wide a range of different levels as possible, there was a gap of only 40 units between the low and medium word lists and a gap of only 47 imagery units between the medium and high word lists.

Future research should be wary of some of the pitfalls and limitations already mentioned. For example, with regards to the word lists, unless more words with imagery ratings become available, it may be best to use two more widely separated word lists as this would eliminate the less than optimal gaps between the imagery groupings and allow the removal of novel words. Another option would be to reduce the number of words in each programme.

Experiment 1

ANOVAs between response bias and imagery level, where imagery level was taken as an indication of sensitivity, were not significant. However, when the group data from Experiment 1 were graphed (Figures 5 and 6) a tendency for a negative relationship was noticed in the data, and was later confirmed in some of the correlations.

With respect to this relationship, the lack of statistical significance was not totally unexpected, as the relationship between these two variables was thought to be a tendency rather than a rule, and was thus likely to be obscured by variations in the sensitivity of individual participants. On the other hand, the significant correlations are interesting as they suggest that, when participants respond according to their natural inclination, as the recognition task becomes increasingly difficult responding becomes more conservative. This lends support to the theory that sensitivity may have an influence on response bias. Because the ANOVAs used an indirect measure of sensitivity as the independent variable (imagery level), whereas individual sensitivity measures were used in the correlations the latter provide an important indicator of a relationship between bias and sensitivity.

Response Bias

From Experiment 1, there was a further finding that was both clear and interesting. That was that participants' natural bias was quite conservative. If participants were unbiased, then the criterion would lie in the middle of the two distributions displayed in

Figure 4, but in the current experiment, participants' criterions were typically positioned to the right of this location. It has long been accepted that it is quite rare for participants to respond without some degree of bias (Balakirshnan, 1997). Rather the positioning of the criterion is influenced by rewards and punishments associated with outcomes (Titus, 1973) which can either be internally based (natural bias) or externally based (e.g., a weighted payoff matrix). As there was no external payoff in force in Experiment 1, the conservative bias may be the result of internal factors.

The internal influences are proposed to be based on social rewards and punishments and these factors can affect the strategy adopted by participants (Hilgendorf & Irving, 1978). For example, when SDT was applied to advertising recognition, it was noted that responding was very liberal. Further investigation lead to the discovery of two biases, social desirability and desire to please the examiner leading to 'yea saying' (Singh, Gilbert, & Churchill, 1986). It is possible that biases such as these were influencing responding in the present study. The results of Experiment 1 indicate that this group would prefer to have fewer hits and fewer false alarms at the expense of more misses and more correct rejections. Therefore, in terms of being incorrect, missing a previously presented word was not perceived to be as undesirable as incorrectly labeling a new word as having been presented previously.

To a degree, conservative bias on the part of normal individuals is in keeping with previous research (Snodgrass & Corwin, 1988; Batters et al., 1985; Branconnier et al., 1982, cited in Snodgrass & Corwin, 1988) where normal participants responded in a slightly conservative manner, while amnesics or demented participants were quite liberal in their responding.

Inaccuracies in measuring bias may have arisen from flaws within the methodology of this experiment. For example, there was considerable variation in the time over which participants completed the trial blocks. While the number of blocks that could be completed in a certain time were restricted to no more than two a day, at least two hours apart, no time constraints were set on how many days could elapse between trial blocks. Some participants completed the trial blocks in a regular manner, undertaking a block each day (except weekends) while other participants were not so consistent. For example, some participants completed more than one trial a day. By contrast, due to external commitments, many found it necessary to take breaks from the testing. While the effect of variable time spacing between trial blocks on criterion stability is unknown, it is possible that when participants resumed testing, a slightly different criterion could have been adopted

Experiment 2

The aim of Experiment 2 was to establish if the relationship noted between bias and sensitivity when responding 'naturally' occurred when bias was imposed. To impose bias, a weighted payoff matrix was introduced with an incentive to accumulate the most points. The same points system that was used in Snodgrass and Corwin's (1988) study to produce either liberal or conservative responding was used in this research.

When the biases from Experiment 1 were compared with those from Experiment 2 (which promoted liberal and conservative responding via two types of points allocation), the biases did not differ significantly. It appears that, given to the participants' tendency towards a naturally conservative bias, it was difficult to make the bias markedly more conservative. The liberal and conservative payoffs, however, produced bias measures that differed significantly from each other at each of the three imagery levels.

The ANOVAs used to investigate any link between sensitivity and imposed bias did not produce significant results for either liberal or conservative responding. This result was echoed in the correlations which produced only one significant correlation out of 12. It is assumed that the trend that presented itself in Experiment 1 was lost by asking participants to adopt particular biases and their desire to comply with the experimenter's instructions.

Maximum likelihood fits

The most disappointing outcome of Experiment 2 was the failure to obtain suitable data from which maximum likelihood fits of the normal-normal SDT model could be derived, and for this reason the single-point estimates of sensitivity were the only option for analysis. However, the use of a rating procedure to provide data for such fits would have made it difficult to specify and compare bias sensibly.

Conclusion

Overall, this research indicated that response bias tended to become more conservative as sensitivity decreased via the manipulation of the imagery values of words in a

memory list. It was also evident that individuals' natural response biases in a Yes/No task were somewhat conservative. Experiment 2 revealed that when response bias was manipulated experimentally it was independent of sensitivity, presumably reflecting the observer's ability to put aside his or her inherent bias when given specific instructions to act in a particular manner.

If a relationship between sensitivity and bias does exist, this would have an impact on other research that has used SDT as an analysis tool to categorise response bias. For example, SDT has been applied to facial identification studies (Shapiro & Penrod, 1986), also referred to as person recognition tasks (Hilgendorf & Irving, 1978) with the purpose of investigating the validity of eye-witness identification. In this research it was noted that many factors could affect sensitivity, for example, race of the suspect, age, exposure time and same verses cross race identification (Shapiro & Penrod, 1986). If these factors have an impact on sensitivity, they could potentially also have an impact on the response bias of the witness.

As this experiment was conducted with only normal participants, the results can not be generalised to other populations. As speculated at the close of Chapter 3, it is possible that the negative trend noticed between response bias and sensitivity may be not apply. The only way to obtain insight into how sensitivity affects amnesic or demented participants' response bias, is to conduct a similar study as the current research using these populations as the volunteers. This type of research would give more insight as to how different populations respond under comparable levels of difficulty. Preliminary testing would be needed to adjust the word lists, using the words with high imagery levels and adjusting the memory word list length. Past research indicates that this may be more difficult due to the low d′ scores that are obtained by amnesics, with most scores being close to zero. It is therefore suggested that different tasks may be more productive, for example, testing memory using pictures or using priming tasks.

The current study has provided some interesting insight into the possibility that response behaviour is affected by sensitivity and suggests that the question deserves further experimental investigation within the field of memory, and across other modalities.

References

Baddeley, A. (1994). Working memory: The interface between memory and cognition. In D.L. Schacter & E. Tulving (Eds.), *Memory Systems 1994,* (pp. 351-368). London: A Bradford Book.

Balakrishnan, J.D. (1997). Form and objective of the decision rule in absolute identification. *Perception and Psychophysics, 59,* 1049-1058

Banks, W.P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74,* 81-99.

Bartus, R.T., Dean, R.L., Beer, B. & Lippa, A.S. (1982). The cholinergic hypothesis of geriatric memory dysfunction. *Science, 217,* 408-417.

Bernback, H.A. (1967). Decision processes in memory. *Psychological Review, 74,* 462-480.

Boles, D.B. (1989). Word attributes and lateralization revisited. *Memory and Cognition, 17,* 106-114.

Brown, A.S. (1976). Catalog of scaled verbal material. *Memory and Cognition, 4,* 1-5.

Brown, J. (1976). An analysis of recognition and recall and of problems in their comparison. In J. Brown (Ed.), *Recall and recognition* (pp.1-35). London: Jonh Wiley & Sons

Brown, J., Lewis, V.J. & Monk, A.F. (1977). Memorability, word frequency and negative recognition. *Quarterly Journal of Experimental Psychology, 29,* 461-473.

Butters, N., Wolfe, J., Martone, M., Granholm, E. & Cermak, L.S. (1985). Memory disorders associated with Huntington's disease: Verbal recall, verbal recognition and procedural memory. *Neuropsychologia, 23,* 729-743.

Carlson, N.R. (1993). *Psychology: The science of behavior.* (4th ed.). Boston: Allyn and Bacon.

Christian, J., Buckley, W., Tarka, M. & Clayton, K. (1978). Measures of free recall of 900 nouns: Correlations with imagery, concreteness, meaningfulness, and frequency. *Memory and Cognition 6,* 379-390.

Cohen, N.J. & Squire, L.R. (1980). Preserved learning and retention of pattern analyzing skill in amnesics: Dissociation of knowing how and knowing that. *Science, 210,* 207-210.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33,* 497-505.

Curry, R.E., Nagel, D.C., & Gai, E.G. (1977). Decision behavior with changing signal strength. *Journal of Mathematical Psychology, 15,* 46-69.

Davelaar, E. & Besner, D. (1988). Word identification: Imageability, semantics and the content function distinction. *Quarterly Journal of Experimental Psychology, 40,* 789-799.

Devane, J.R. (1988). Sensitivity to the imagery values of words. *Perceptual and Motor Skills, 66,* 300-302.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121,* 275-277.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory and Cognition, 24,* 523-533.

Dorfman, D.D. & Alf, E. Jr. (1968). Maximum likelihood estimates of parameters of signal-detection theory - a direct solution. *Psychometrika, 33,* 117-124.

Egan, J.P. (1975). *Signal detection theory and ROC analysis.* New York: Academic Press.

Eichenbaum, H. (1994). The hippocampal system and declarative memory in humans and animals: Experimental analysis and historical origins. In D.L. Schacter & E. Tulving (Eds.), *Memory Systems 1994,*(pp. 147-201). London: A Bradford Book.

Estes, W.K. & Maddox, W.T. (1995). Interactions of stimulus attributes, base rates and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 1075-1095.

Ferlazzo, F., Conte, S. & Gentilomo, A. (1993). Event-related potentials and recognition memory: the effects of word imagery value. *International Journal of Psychophysiology, 15,* 115-122.

Gardner, R.M. (1997). Misconceptions about classical psychophysics and the measurement of response bias. *Perceptual and Motor Skills, 84,* 587-594.

Gardiner, J.M & Gregg, V.H. (1997). Recognition memory with little or no remembering: Implications for a detection model. *Psychonomic Bulletin and Review, 4,* 474-479.

Glanzer, M & Adams, J.K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5-16.

Green, D.M. (1964). General prediction relating yes-no and forced-choice results. *Journal of the Acoustical Society of America, 36*, 1042.

Green, D.M & Swets, J.A. (1966). *Signal detection theory and psychophysics.* New York: Wiley. Reprinted 1974 by Krieger, Huntington, NY.

Hannay, H.J. (1986). *Psychological measurement techniques and their application to neuropsychology.* Oxford University Press: Oxford.

Hilgendorf, E.L. & Irving, B.C. (1978). Decision criteria in personal recognition. *Human Relations, 31,* 781-789.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list length paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 302-313.

Hirshman, E. & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science, 9,* 61-65.

Johnson, M.K. & Chalfonte, B.L. (1994). Binding complex memories: The role of reactivation and the hippocampus. In D.L. Schacter & E. Tulving (Eds.), *Memory Systems 1994,* (pp. 1-38). London: A Bradford Book.

Kaschak, M.P. & Charnetski, C.J. (1998). Nonsense syllables in associative recognition: Implications for global memory tasks. *Psychological Reports, 82,* 95-105.

Lockhart, R.S. & Murdock, B.B. (1970). Memory and the theory of signal detection. *Psychological Bulletin, 74,* 100-109.

Luce, R.D. (1963). Detection and recognition. In R.D. Luce, R.R. Bush & E. Galanter (Eds.). *Handbook of Mathematical Psychology*, New York : Wiley.

Macmillan, N.A. & Creelman, C.D. (1990). Response bias: Characteristics of detection theory, threshold theory and "nonparametric" index. *Psychological Bulletin, 107,* 401-413.

Macmillan, N.A. & Creelman, C.D. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.

Macmillan, N.A. & Creelman, C.D. (1996). Triangles in ROC space: History and theory of "nonparametric" measures of sensitivity and response bias. *Psychonomic Bulletin and Review,3,* 164-170.

Maddox, W, T. & Estes, W.K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 539-559.

McNicol, D. (1972). *A primer of signal detection theory.* London: George Allen & Unwin Ltd.

Marks, D.F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64,* 17-24.

Matlin, M.W. (1994). *Cognition.* (3rd ed.). New York: Harcourt Brace Publishers.

Noldy, N.E., Stelmack, R.M. & Campbell, K.B. (1990). Event-related potentials and recognition memory for pictures and words: The effects of intentional and incidental learning. *Psychophysiology, 27,* 417-428.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart, & Winston.

Paivio, A. (1976). Imagery in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp.103-129). London: John Wiley & Sons

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford: Oxford University Press.

Paivio, A., Yuille, J.C. & Madigan, S.A. (1968). Concreteness, imagery and meaningfulness for 925 nouns. *Journal of Experimental Psychology Monograph Supplement, 76,* 1-25.

Peterson, M.J. & Graham, S.E. (1974). Visual detection and visual imagery. *Journal of Experimental Psychology, 103,* 509-514.

Pollack, I. & Norman, D.A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science, 1,* 125-126.

Reed, J.M. & Squire, L.R. (1997). Impaired recognition in patients with lesions limited to the hippocampal formation. *Behavioral Neuroscience, 111,*667-675.

Rehm, L.P. (1973). Relationships among measures of visual imagery. *Behaviour Responses & Therapy, 11,* 265-270.

Reinitz, M.T., Verfaellie, M. & Milberg, W.P. (1996). Memory conjunction errors in normal and amnesic subjects. *Journal of Memory and Language, 35,* 286-299.

SAS. 1989. SAS/STAT User's Guide, Version 6, 4th Ed. SAS Institute, Inc., Cary, NC.

Schacter, D.L. & Tulving, E. (1994). What are the memory systems of 1994? In D.L. Schacter & E. Tulving (Eds.), *Memory Systems 1994,* (pp. 1-38). London: A Bradford Book.

See, J.E., Warm, J.S., Dember, W.N. & Howe, S.R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors, 39,* 14-29.

Shapiro, M.L. & Olton, D.S. (1994). Hippocampal function and interference. In D.L. Schacter & E. Tulving (Eds.), *Memory Systems 1994,* (pp. 87-118). London: A Bradford Book.

Shapiro, P.N. & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139-156.

Sheehan, P. (1967). Shortened form of Betts' Questionnaire Upon Mental Imagery. *Journal of Clinical Psychology, 23,* 386-389.

Shriffrin, R.M. & Steyvers, M. (1997). A model for recognition memory: REM - retrieving effectively from memory. *Psychonomic Bulletin and Review, 4,* 145-166.

Singh, S.N & Churchill, G.A. (1988). Using the theory of signal detection to improve ad recognition tests. *Journal of Marketing Research, 23,* 327-336.

Snodgrass, J.G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34-50.

Squire, L.R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D.L. Schacter & E. Tulving (Eds.), *Memory Systems 1994,* (pp. 203-232). London: A Bradford Book.

Stillman, J.A., Brown, G.M. & Troscianko, T. (1998). Influence of sensitivity on response bias in taste and audition. Unpublished manuscript.

Stubbs, D.A. (1976). Response bias and the discrimination of stimulus duration. *Journal of the Experimental Analysis of Behavior, 25,* 243-250.

Swets, J.A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99,* 181-198

Tanner, W.P.J. & Swets, J.A. (1954). A decision-making theory of visual detection. *Psychological Review, 61,* 401-409.

Titus, T.G. (1973). Continuous feedback in recognition memory. *Perceptual and Motor Skills, 37,* 771-776.

Warrington, E.K. & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature, 217,* 972-974.

Warrington, E.K. & Weiskrantz, L. (1970). Amnesic syndrome: Consolidataion or retrieval? *Nature, 217,* 628-630.

Warrington, E.K. & Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia, 12,* 419-428.

Wickelgren, W.A. & Norman, D.A. (1966). Strength models and serial position in short term recognition memory. *Journal of Mathematical Psychology, 3,* 316-347.

Wild, C.J. & Seber, G.A.F. (1994). *Introduction to Probability and Statistics.*

APPENDIX A

Information Sheet

You are invited to participate in research conducted towards a Masters of Arts in Psychology. The research is concerned with how individuals decide on which response to make under conditions of uncertainty. To investigate these processes a recognition memory task will be used. This task will be made more or less difficult by changing the imagery level of the words which participants will have to decide was either seen, or not seen, in a list of words previously. Sometimes points will be allocated for correct or incorrect responses to asses the effect of various consequences on response decisions.

The research will consist of an initial session where an explanation and a preface will be given. After this you may begin your trials. This research consists of 27 short (10 minute) trials, which involve responding on a computer keyboard. The work can be completed over a period of weeks to suit you. You may do a maximum of two trials per day, but there must be a minimum of **2 hours** between each trial. The data generated from these trials will be stored on computer disk.

If you agree to take part in this research you will receive more specific instructions on the task you are asked to participate in. You will have the right to withdraw from the project at any time and to ask any questions about the study during participation. It is important to note that participation in this research is completely voluntary and all information obtained will remain confidential. At the completion of the study a summary of the findings will be made available to you.

The primary researchers in this project include myself, Jane Burgess, and my supervisor, Jennifer Stillman. Should you have any further inquirers or concerns about the research, feel free to contact either of us on the below contact numbers:

Jane Burgess
418 1032
Jane_Burgess@xtra.co.nz

Dr. Jennifer Stillman
443 9770
J.Stillman@massey.ac.nz

CONSENT FORM

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I understand I have the right to withdraw from the study at any time and that participation in this study is completely voluntary.

I agree to provide information to the researchers on the understanding that my name will not be used without my permission. *(The information will be used only for this research and publications from this research project).*

I agree to participate in this study under the conditions set out in the Information Sheet.

Signed: ....

Name:
....

Date: ....

Instructions

Firstly, thank you for agreeing to participate in this research. Please remember that your involvement in this study is completely voluntary and that you are free to withdrawal at any time.

Below are the instructions that will direct you in running the programme. It is suggested that you read the instructions through once before starting the programme. It you have any questions or problems feel free to ask the researcher for assistance. Access to the programs is obtained through 'Word Explorer'.

It is estimated that each trail should take around or less than ten minutes. On page 3 is a grid on which the order your particular trials should be run is given.

Once 'Word Explorer' has been accessed, select the appropriate file as instructed on page three and then *double click* on this file using the *left mouse button*. This will automatically run the selected programme and the first screen will appear:

1. ***Toggle Feedback***: - off (default)
2. ***Payoff Matrices***: 1. liberal, 2 neutral, 3 conservative, 4 none.
3. ***Continue***

For this screen you do not need to alter anything, so just press '3' and then *enter* to continue to the next screen. The second screen will contain several prompts asking for information, firstly:

Enter id: JS1

Session number: 1

Male (m) or female (f): F

To complete this section consult your grid on the page 3. The programme will begin immediately after this.

This is a recognition task, so pay attention to the words that appear. A total of 120 will be presented to you. Following this another screen will appear, it states the following:

Get ready for test session.
Press B if you have seen the word before, otherwise, press N.

This will appear on the screen for 18 sec, it is suggested that you place your right pointer finger on the 'N' key and your left pointer finger on the 'B' key. To help remember this the 'N' stands for *New* and the 'B' stands for *Before.* A statement will appear on the screen telling you when the session is going to begin.

At the completion of the trial, a summary will appear.

You will notice on the grid that the first 9 sessions are in the 'none' category. This means that there is no payoff. However, the other sessions have either a liberal or a conservative payoff (explained below). In these sessions points will be allocated according to your response. When the summary is given at the end you will be given a score. There will be a reward given to the people who obtain the overall high score for both of the conditions.

Liberal

Using this payoff it is better for you to make more hits and false alarms than it is for you to make misses or correct rejections. The easiest way to think of this is "if in doubt, say old - press 'B'".

Conservative

The conservative payoff is the exact opposite of the liberal payoff and in this case you should have a stronger tendency to say no. Therefore "if in doubt, press 'N'.

When doing trials 10-27 the summary information should be copied on to the scoring sheets contained in your folder. The scoring sheet should be kept next to you by the computer.

Once you have completed this, using the left mouse button click on the 'X' in the upper right-hand corner to close the programme, this will take you back to 'explorer'.

Once you have completed the session call for the researcher to collect your data, or put it in your folder and you are free to leave.

Disk	*Block*	*File Name*	*Id*	*Session*	*Finished*
1 None	1	Medcor1	ID2	1	
	2	Locor1	ID1	1	
	3	Hicor1	ID3	1	
	4	Locor1	ID1	2	
	5	Hicor1	ID3	2	
	6	Medcor1	ID2	2	
	7	Medcor1	ID2	3	
	8	Hicor1	ID3	3	
	9	Locor1	ID1	3	
2 Conservative	10	Hicor2	ID3	1	
	11	Locor2	ID1	1	
	12	Medcor2	ID2	1	
	13	Locor2	ID1	2	
	14	Medcor2	ID2	2	
	15	Hicor2	ID3	2	
	16	Hicor2	ID3	3	
	17	Medcor2	ID2	3	
	18	Locor2	ID1	3	
3 Liberal	19	Hicor3	ID3	1	
	20	Medcor3	ID2	1	
	21	Locor3	ID1	1	
	22	Medcor3	ID2	2	
	23	Locor3	ID1	2	
	24	Hicor3	ID3	2	
	25	Locor3	ID1	3	
	26	Hicor3	ID3	3	
	27	Medcor3	ID2	3	

APPENDIX B

Conservative and liberal conditions.

For trial blocks 10-27 you will be using the disk marked *conservative* or *liberal*. Under these conditions you will accumulate points which will be contingent on your responses. Before a full explanation is given on how the point system works, firstly an overview of some terms may be useful.

Hit : a hit occurs when the word is old and your keyboard response is 'B'.
False Alarm: a false alarm occurs when the word was new but your keyboard response is 'B'.
Miss: a miss occurs when the word was in fact old, but your keyboard response is 'N', and
Correct Rejection: a correct rejection occurs when the word was new and your keyboard response is 'N'.

Under the *Conservative* condition points will be allocated as follows:
10 points will be given for each *correct rejection*
1 point for each *hit*
-10 points for each *false alarm*, and
-1 point for each *miss.*

On the other hand, for the *Liberal* condition points will be allocated as follows:
10 point will be given for each *hit,*
1 point for each *correct rejection,*
-10 points for each *miss,* and
-1 point for each *false alarm.*

Your aim is to accumulate the largest amount of points. Therefore for trials 10 and onwards, check which condition you are using and try to get as many points as you can. The participant whom obtains the highest score for either of the conditions will get a 'token' gift.

Good luck.

Conservative and liberal conditions.

For trial blocks 10-27 you will be using the disk marked *conservative* or *liberal*. Under these conditions you will accumulate points which will be contingent on your responses. Before a full explanation is given on how the point system works, firstly an overview of some terms may be useful.
Hit : a hit occurs when the word is old and your keyboard response is 'B'.
False Alarm: a false alarm occurs when the word was new but your keyboard response is 'B'.
Miss: a miss occurs when the word was in fact old, but your keyboard response is 'N', and
Correct Rejection: a correct rejection occurs when the word was new and your keyboard response is 'N'.

Under the *Conservative* condition points will be allocated as follows:
10 points will be given for each *correct rejection*
1 point for each *hit*
-10 points for each *false alarm*, and
-1 point for each *miss.*

On the other hand, for the *Liberal* condition points will be allocated as follows:
10 point will be given for each *hit,*
1 point for each *correct rejection,*
-10 points for each *miss,* and
-1 point for each *false alarm.*

Your aim is to accumulate the largest amount of points. From your data we can measure both your response bias and your sensitivity to the differences between previously seen and new words. Theses are independent factors. If you remain focused on the task your sensitivity should either remain the same throughout the study, or increase a little as a result of practice. In order to ensure you concentration, where points are given there will be a 25% penalty if there is a decrease of 10% or more in your sensitivity compared to the initial phase of the study. Whatever the payoff, the best strategy to adopt is to respond a honestly as you can, but when you are uncertain as to the correct response, bear in mind the appropriate payoffs for the current block of trials.
Good luck.

Scoring Card

Id: _________ *Session Number :*________

'Hit' score: ____

'False alarm' score: ____

P(hits): ____

P(false alarms): ____

'Correct rejection' score: ________________________________

'Miss' score: ____

Points accumulated: ____

Scoring Card

Id: ______ *Session Number:*______

'Hit' score: ____

'False alarm' score: ____

P(hits): ____

P(false alarms): ____

'Correct rejection' score: ________________________________

'Miss' score: ____

Points accumulated: ____

Appendix C

				Participant Number				
Measures	1	2	3	4	5	6	7	8
Low Imagery								
D'	2.08	1.149	1.658	0.896	2.028	1.375	1.592	1.891
C	0.64	0.393	0.256	0.434	0.268	0.66	-0.086	0.022
A'	0.869	0.799	0.869	0.754	0.904	0.822	0.864	0.896
B''_D	0.799	0.426	0.345	0.438	0.289	-0.392	-0.077	0.03
Medium Imagery								
D'	2.1	1.534	1.751	1.085	2.765	1.617	1.927	2.456
C	1.078	0.397	0.092	0.543	0.001	0.836	0.122	0.192
A'	0.862	0.851	0.882	0.785	0.955	0.84	0.899	0.936
B''_D	1.122	0.484	0.133	0.772	0.001	0.923	0.122	0.088
High Imagery								
D'	2.194	1.934	2.229	1.406	2.263	1.916	2.654	3.372
C	0.604	0.626	0.387	0.843	-0.118	0.588	0.266	-0.141
A'	0.902	0.881	0.916	0.818	0.925	0.882	0.946	0.975
B''_D	0.591	0.764	0.569	1.578	-0.16	0.377	0.223	-0.507

Table C.1 Table of measures for individual results for Experiment 1

Measures	Participant Number 1	2	3	4	5	6	7	8
Low Imagery								
D'	2.514	1.147	1.579	0.903	1.94	1.674	1.446	1.837
C	0.244	-0.33	-0.2	0.634	-0.788	0.583	-0.337	-0.394
A'	0.939	0.802	0.861	0.753	0.872	0.859	0.842	0.883
B''_D	0.201	-0.036	-0.269	0.64	-0.851	-0.346	-0.302	-0.534
Medium Imagery								
D'	2.354	1.526	2.147	1.704	1.997	1.7	1.996	3.018
C	-0.52	-0.14	0.037	1.062	-0.76	0.342	-0.252	-0.317
A'	0.918	0.856	0.918	0.833	0.878	0.872	0.902	0.96
B''_D	-0.541	-0.171	0.054	1.512	-0.927	0.377	-0.253	-0.145
High Imagery								
D'	4.116	1.569	1.98	1.817	2.08	2.13	2.805	2.727
C	-0.413	0.057	0.023	1.006	-0.787	0.355	-0.239	-0.227
A'	0.986	0.862	0.904	0.846	0.883	0.91	0.954	0.95
B''_D	-0.404	0.069	0.034	1.884	-1.061	0.228	-0.199	-0.819

Table C.2 Table of measures for individuals under the liberal payoff matrix

	Participant Number							
Measures	1	2	3	4	5	6	7	8
Low Imagery								
D'	2.554	1.276	2.219	0.906	2.119	1.558	1.34	2.281
C	1.009	0.526	0.724	0.93	0.494	0.569	0.113	0.08
A	0.896	0.814	0.897	0.748	0.909	0.848	0.832	0.927
B''_D	0.831	0.571	0.974	0.939	0.533	-0.338	0.101	0.109
Medium Imagery								
D'	5.304	1.081	2.302	1.651	2.685	2.024	2.034	2.802
C	1.515	0.68	0.494	1.303	0.786	0.998	0.234	-0.018
A	0.968	0.781	0.916	0.813	0.921	0.863	0.906	0.956
B''_D	1.577	0.83	0.714	1.854	0.959	1.102	0.234	-0.008
High Imagery								
D'	3.299	2.002	2.407	2.343	2.955	1.964	2.566	2.917
C	0.637	1.127	0.342	1.368	0.532	0.611	0.481	0.001
A	0.957	0.851	0.929	0.851	0.949	0.884	0.933	0.961
B''_D	0.623	1.375	0.503	2.561	0.718	0.393	0.401	0.003

Table C.3. Table of measures for individuals under the conservative payoff matrix

	Participant Number							
Measures	1	2	3	4	5	6	7	8
Low Imagery								
AA	3.1816	1.0508	1.2122	0.8796	1.9024	-5.827	1.6074	1.5248
B	1.4282	0.844	0.487	0.9802	0.8524	-4.369	1.2373	0.4764
d'e	2.6205	1.1397	1.6304	0.8884	2.054	3.4596	1.4369	2.0656
Medium Imagery								
AA	2.5949	1.9994	1.2968	0.4667	2.0631	1.4634	1.9792	5.9353
B	0.9212	1.6184	0.3841	0.4053	0.6392	0.8115	0.9946	3.3837
d'e	2.7013	1.5272	1.8739	0.6642	2.5172	1.6157	1.9846	2.7079
High Imagery								
AA	3.3076	1.2927	1.3363	0.0082	1.8535	3.6956	3.2707	0.9991
B	1.0435	0.6391	0.3585	0.0681	0.4821	2.1137	1.3942	-0.446
d'e	3.2372	1.5773	1.9673	0.0154	2.5012	2.3738	2.7322	3.6082

Table C.4 Table of results for individual data for slope, intercept and d'_e