Effects of Sample-set size on Delayed-matching-to-sample

A Thesis submitted
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

At the University of Waikato

By

Coreen Adamson

University of Waikato
January 2002

In a series of experiments the discriminative performance of university students was assessed using a Delayed-Matching-to-Sample task. White disks, presented on computer screens, served as stimuli. For the general DMTS task, used here, a sample stimulus was presented to the participant. After a variable delay (0.05, 4, 8 or 16 s) two comparison stimuli were presented on the computer screen. One of these comparison stimuli was the same size as the sample stimulus (matching comparison) and the other was slightly larger or smaller than the sample stimulus (non-matching comparison). The sample stimuli were not intended to appear as comparison stimuli and for each of the sample stimuli there were two possible non-matching comparison sample stimuli, one larger and one smaller than the sample stimulus. The first three experiments showed that it was possible to use a titration procedure to get similar levels of performance across participants, at zero delay. During the titration procedure the size of the non-matching comparison stimuli was changed, for each participant until they were 100% correct at 0.05 s delay. The first two of these experiments also showed that there were generally only small decreases in performance as the delay between the presentation of the sample stimuli and the comparison stimuli was increased. The fourth experiment showed that performance on the DMTS task did not change when feedback was given for correct responses. The fifth and sixth experiments showed that as the number of sample stimuli used in the DMTS task, was increased there was a systematic decrease in *a*, a measure of discrimination at zero delay, but no systematic change in *b*, a measure of the rate of decrement in performance with increasing delay. It was not clear whether this decrease in *a* was a product of the increase in the sample-set size or of the associated decrease in the correspondence of sample stimuli on consecutive trial pairs. Experiment 7 suggested that, when the number of sample stimuli used for the DMTS was held constant and the ratio of corresponding to non-corresponding trial pairs was varied, there was no systematic change in either *a* or *b*. However, the results of this experiment were confounded by a ceiling effect and so a different procedure was required. The procedure used for the remaining experiments was a two-sample DMTS task, where the sample stimuli were also used as comparison stimuli. This type of

DMTS task is similar to that typically used in discrimination tasks with animal subjects. The eighth and ninth experiments were designed to investigate whether this type of procedure could be successfully used with human participants and found that it could be. Experiment 10 also used this type of two-sample DMTS task to investigate the effect of decreasing the ratio of corresponding to non-corresponding trial pairs. It was found that, when there was no ceiling effect, varying the ratio of corresponding to non-corresponding trial pairs did not result in a systematic change in either *a* or *b* as in Experiment 7. These results are discussed in terms of the effects of range, sample-set size and of the ratio of corresponding to non-corresponding trial pairs on human performance on DMTS.

Acknowledgements

My heartfelt thanks to the following people:

To James, Mary and to Bill: for their excellent supervision and for hanging in there with me.

To my family and to Ian: thank you for your patience, your understanding, your support and for believing in me.

To all my friends: thanks for all your friendship and your support over the years.

To Rob Bakker: for the friendly face and the excellent programming.

Finally to all those fellow Doctoral and Masters students who provided a friendly ear and lots of support over the years.

Table of contents

Appendices C to L are recorded on a CD Rom, inside the back cover of the thesis.

List of Figures

List of Tables

List of Appendices

The study of human memory has involved the development and use of a number of different experimental methodologies. One method that has been commonly used to study memory function in patients with dementia's such as Alzheimer's disease is the Peterson and Peterson task (1959). They developed a task that was designed to investigate recall across a range of short delay intervals between the presentation of a sample and the recall test. A typical trial began with the presentation of three consonants (stimuli) which a patient was required to remember after delays of up to 30 s. During the delay interval the participant was required to perform a distracter task, designed to prevent verbal repetition or rehearsal of the stimuli. Once the delay interval had finished a recall test was administered and the participants were typically required to produce the stimuli verbally.

The most common finding of the Peterson and Peterson task (1959) is that patients with Alzheimer's disease have impaired performance on this task, in comparison to an age-matched control group (Morris, 1992). However, it has been suggested that the Peterson and Peterson task (1959) gives a measure of memory which is confounded by interference from the distracter task (Morris, 1992). A non-verbal version of the Peterson and Peterson task (1959) was developed by Sullivan, Corkin and Growden (1986, cited in Morris, 1992) where the stimuli were a series of taps on corsi, or small wooden blocks. As these stimuli could not be verbally rehearsed it was not necessary to include a distracter task during the delay interval. Despite the removal of the distracter task the performance of patients with Alzheimer's disease was still found to be impaired. Traditionally this impairment in accuracy was seen as an increase in the rate at which information was forgotten by patients with Alzheimer's disease (Kopelman, 1992).

Another task which has commonly been used to investigate human memory is the Huppert and Piercy (1978) task (Kopelman, 1985; Martone, Butters & Trauner (1986); Hart, Kwentus, Taylor, & Harkins (1987) and Wixted & Ebbesen (1991)) Huppert and Piercy (1978) developed this recognition task which did not use a distracter task during the delay interval. This means that the measure of memory it provides is not confounded by the distracter task. In a

typical Huppert and Piercy (1978) task participants were presented with 120 slides of coloured pictures (stimuli) and asked to remember them. A memory test was administered 10 minutes, 1 day and 7 days later. This memory test involved presenting participants with a sample of 40 of the original pictures, randomly interspersed with 40 new coloured pictures (distracters). The participants were required to say whether they had or had not seen each of the pictures presented during the memory test. Both Huppert and Piercy (1978) and Kopelman (1992) found that when patients with dementia were given longer exposure times, to the stimuli, that their rate of forgetting did not differ from that of a control group. They suggested, based on these results that the memory deficit evident with dementia patients was due to a deficit in initial acquisition rather than an increased rate of forgetting over time.

The delayed matching-to-sample task (DMTS) is a methodology, which has also been used frequently to study memory (White & Alsop, 1993). This procedure is based on a matching-to-sample task (MTS) which requires that a subject learns to make a particular response in the presence of a stimulus. The MTS task involves the presentation of a sample stimulus and two comparison stimuli, where one of the two comparison stimuli is the same as the sample stimulus. The comparison stimuli may be presented concurrently with the sample stimulus, which is termed simultaneous matching to sample, or the comparison stimuli may be presented immediately following the sample stimulus, which is often termed, zero-delay matching to sample. Control by a sample stimulus is demonstrated when the subject responds to the comparison stimulus that matches the sample stimulus. When used to study memory, a delay is included between the presentation of the sample stimulus and the comparison stimuli, and this is termed delayed matching to sample (DMTS). The most common finding with the DMTS task is that, as the duration of the delay between the presentation of the sample stimulus and the comparison stimuli increases, matching accuracy decreases (e.g., Wixted & Dougherty, 1996).

The DMTS task has been used mainly with animals although there are some studies using human participants. Studies of animal and human memory using the DMTS task differ notably in the type of sample stimuli they typically

use. Many of the stimuli used for the DMTS tasks with animals (such as red and green stimuli) are considered inappropriate with humans as the stimuli could be labelled, and hence rehearsed, during the delay. In overcoming this problem, many studies with human participants have used more than two sample stimuli. Money, Kirk and McNaughton (1992), for example, used a DMTS task to study memory in patients with Alzheimer's disease. They report successfully using a DMTS procedure in which they presented patients with one of 13 different sized circles and asked them to choose the size-matched comparison stimulus after delays ranging from 0.5 to 32 s. Parr (1992) also successfully used a DMTS task with eight different sized squares to study human visuospatial memory. Participants were presented with one of the different sized squares and after delays from 0 to 15 s were required to decide whether a single comparison stimulus was the same size as the sample stimulus or not. Holdstock, Shaw and Aggleton (1995) also used a DMTS task, where the stimuli presented on each trial were different or trial-unique abstract patterns, to study the performance of amnesic participants successfully. The use of more than two sample stimuli presents a potential problem as it is possible that the sample-set size, or the number of distinctly different sample stimuli, used in a DMTS task may have an effect on accuracy.

Adamson (1995) used a DMTS procedure, similar to Money et al.'s. (1992) with university students and her results suggested that the number of stimuli in the sample set influenced accuracy. She used filled white disks, presented on a computer screen, as sample stimuli. As the number of sample stimuli was increased across conditions, from one to four, she found that matching accuracy decreased.

Several other studies, using animals as subjects, have examined the effect of varying the number of sample stimuli used in a DMTS procedure. Etkin and D'Amato (1969) investigated the impact of varying the sample-set size in a DMTS task with monkeys. They used two, three or four sample stimuli and delays ranging from 1 to 18 s. Accuracy for all three sample sets was found to decrease as the delay interval was increased but the size of the sample set was found to have no significant effect. Mason and Wilson (1974) also studied the

impact of varying the number of sample stimuli, using four or six samples, in a DMTS task with monkeys. Their results were consistent with those of Etkin and D'Amato (1969) in that they found accuracy did not change as the size of the sample set was increased.

Worsham (1975) conducted a DMTS study with monkeys using two, three and seven sample stimuli in each set, at either a 60 s or a 2 s delay. He found, as he predicted, that accuracy increased as the size of the sample set increased. This finding is consistent with that of Mishkin and Delacour (1975) who compared the performance of monkeys on a DMTS task with either trial-unique (thus, multiple) stimuli or a single pair of objects, at a single delay of 10 s. They found that the trial-unique group acquired the task, to a criterion of 90%, much faster than the single-pair group.

Roberts (1980) also examined the effect of increasing the number of sample stimuli in a DMTS procedure. Roberts (1980) predicted that as the sample-set size was increased, from one to two, accuracy would decrease. He investigated the matching accuracy of pigeons with both a single sample stimulus and two sample stimuli. He points out that the single-sample condition provided a useful comparison condition even though it differed from the typical two sample DMTS in that there was no conditional discrimination required. In contrast to Worsham (1975), he found, as he predicted, that when there were two sample stimuli, matching was less accurate than when there was a single sample stimulus. This finding is consistent with that of Adamson (1995), who found that accuracy decreased as the number of stimuli in the sample set was increased from one to four.

These studies provide three inconsistent results. One, that sample-set size had no effect on matching accuracy (Etkin and D'Amato, 1969; Mason and Wilson, 1974). Two, that matching accuracy increased as the sample-set size was increased (Worsham, 1975; Mishkin and Delacour, 1975). Three, that accuracy decreased as sample-set size was increased (Roberts, 1980; Adamson, 1995). Given this inconsistency it is not possible to draw any singular conclusion about the impact of the number of samples on matching accuracy. Each of the studies examined a comparatively small and dissimilar range of sample-set sizes and none

provides a parametric investigation of the effect of sample-set size on discrimination performance. However, these studies do indicate that the number of sample stimuli used in a DMTS task may have an impact on matching accuracy.

With the exception of Adamson (1995) all of these studies used percent correct to compare the effect of sample-set size on matching accuracy across a small number of delays. Nevin and Grosch (1990) pointed out that percent correct is a measure of accuracy which has an upper bound of 100%, or perfect performance, and an effective lower bound of 50%, or chance responding. When accuracy is near 0 or 100% changes in performance are constrained to be small in comparison to changes in performance which occur when percent correct is at a level such as 75%. Nevin and Grosch (1990) suggest that this can make the comparison of changes in performance across different levels of accuracy ambiguous. Another disadvantage of using percent correct is that it is a measure of performance which is confounded by response bias (Money, Kirk and McNaughton, 1992; Ruske, Fisher and White, 1997; White 2001).

White and McKenzie (1982) suggested that in a typical memory task the (animal) subject is presented with one of two discriminative stimuli and after a varying delay interval is required to report which of the two stimuli had been presented. Thus, remembering can be viewed as a discriminative behaviour that is under delayed stimulus control, where the main effect of the delay interval is to attenuate the control, which is exerted by the discriminative stimulus (White 1985). White and McKenzie (1982) suggested that if remembering is viewed as a discriminative behaviour then log *d*, a measure of discriminative performance derived by Davison and Tustin (1978) (see Appendix A) for the derivation of log *d*), could be used to quantify the extent to which discriminability decreases as the delay interval is increased in DMTS. White and Wixted (1999) reported that log *d* provides a good measure of discriminative performance in DMTS.

White and McKenzie (1982) showed that a negative exponential function could be used to quantify the relation between discrimination at zero delay as measured by the Davison and Tustin (1978) model and the rate at which

discriminability decreased with increasing delay. This function is:

$$\log d_t = \log d_0 \exp(-bt), \qquad (1).$$

where $\log d_0$ is a measure of initial discriminability or performance at zero delay and b is a constant which describes the rate at which $\log d_t$ decreases over time. Therefore, b can be seen as providing a measure of the extent to which accuracy decreases as the delay interval (t) is increased and so quantifies the extent to which discrimination becomes more difficult as the delay increases.

White (1985) examined whether the rate (b) at which discriminability ($\log d_t$) decreased as the delay interval increased was dependent on the initial discriminability ($\log d_0$) of the sample stimulus. He compared the matching performance of pigeons on a DMTS task with small and large wavelength differences between sample stimuli. He found that $\log d_0$ increased as the physical disparity between the sample stimuli was increased, while b did not change. White (1985) argued that his data showed that the two measures of matching performance provided by the negative exponential fit, the measure of initial stimulus discriminability ($\log d_0$) and the rate at which discriminability decreases with increasing delays (b), are independent.

Using a negative exponential function to quantify the decrement in discriminability as the delay interval is increased relies on the assumption that the decrement which occurs is constant (White 1985, 1991). This assumption is based on the theory that remembering is direct. White (1985, 1991) suggested that the decrement in discriminability which occurs as the delay interval is increased may be analogous to the decrement in discriminability which occurs as spatial distance increases. As the spatial distance of a stimulus is increased it becomes less discriminable and in the same way, when the temporal distance of a stimulus is increased a decrement in discriminability might also be expected. This theory of memory or remembering does not distinguish between the encoding of a stimulus and the remembering of a stimulus. Thus, White (1991) suggested that a function which quantified the decrement of discriminability which occurs, with increasing delay, should decrease at a constant rate.

Wixted and Ebbesen (1991) compared how well a number of different functions; including an exponential, an exponential-power, a hyperbolic, a power, a linear, and a logarithmic function, described the rate of decay from a series of experiments. They suggested that while an exponential function did provide a good description of the rate of decay, that the power function provided the most accurate description of the rate of decay, in terms of goodness of fit. The disadvantage of using a power function to describe changes in discriminability with increasing delay is that the power function is undefined when the delay interval is equal to zero (White, 2001; Wixted and Ebbesen, 1991; White, Ruske and Colombo, 1996). Thus, the power function is not able to quantify discriminative performance in the same way as the negative exponential function and provide independent measures of initial discriminability (log d_0) and the rate of decay (b). White (2001) and White et al. (1996) suggested a function which has similar properties to a power function but still allows data to be described in terms of initial discriminability and rate of decay in the same manner as for the negative exponential.

This function is:

$$\log d_t = \log d_0 \exp\left(-b\sqrt{t}\right) \qquad (2).$$

where log d_0, b and log d_t are the same as in Equation 1.

There is a potential problem with using log d to analyse data from studies of human memory. As mentioned earlier many studies with human participants have used tasks involving more than two sample stimuli. Parr (1992), for example, used eight different sized squares. Money, Kirk and McNaughton (1992) used filled circles of 13 different sizes as sample stimuli and Holdstock et al (1995) used trial unique stimuli. Using more than two sample stimuli presents several potential problems. One, it is not possible to calculate log d and two, is that the sample-set size, or the number of distinctly different sample stimuli, used in a DMTS task may have an effect on accuracy.

It is not possible to calculate log d when more than two sample stimuli are used. However, it is possible to calculate logit p an estimate of log d (Nevin and Grosch, 1990). This function is:

$$\text{logit } p = \log (p/(1-p)) \tag{3}$$

where p is the number of correct responses. Logit p, as a measure of accuracy, has the advantage of not being bounded by both 0 and 100%, as is percent correct. Logit p is bounded by 0 when accuracy is at 50%, or at chance level, but has no upper bound (Nevin and Grosch, 1990). The problem with this measure of accuracy is that it confounds discriminability and response bias, as does percent correct. This means that logit p is only equal to log d when there is no bias. However, White (1985) and White (2001) argued that, when data are averaged across participants, it is reasonably safe to assume that little or no systematic response bias exists in the averaged data and that logit p is equal to log d. Money et al. (1992) used logit p as an estimate of log d and fitted negative exponential functions (Equation 1) to their data. They found that accuracy decreased with increasing delay and that the data were well described by Equation 1. Here, the parameter log d_0 will be referred to as a. Money et al. (1992) reported that a differed across their two groups, but that the b values were the same. This finding provides support for White's (1985) suggestion that a (log d_0) and b are independent.

As noted previously, the DMTS procedure is often used in research with human subjects to assess matching accuracy and the number of sample stimuli used varies across these studies. Hence, it is important to establish what impact varying the number of sample stimuli used in a DMTS task has on the matching accuracy of humans. Thus, the intention was to examine parametrically, the effect that the number of sample stimuli has on matching accuracy in humans using a procedure similar to that used Money et al. (1992) and Adamson (1995). A range of sample-set sizes which encompassed the number of samples used in the three animal studies mentioned above was used. However, one potential limitation of the research by Money et al. (1992) and Parr (1992) is that stimuli of different

sizes might have an influence on performance independent of the proportional difference between the sample and the comparison stimuli. Adamson (1995) showed, using four white disks measuring 2, 4, 6 and 8 cm in diameter, as sample stimuli, that the absolute disk size had no systematic effects on accuracy. However, Adamson's (1995) results also showed a great deal of inter-participant variability. She found this variability problematic when interpreting the data. This issue will be discussed in Experiment 1.

Experiment 1

As mentioned previously, the DMTS procedure has often been used to examine human memory. However, as Adamson (1995) noted the results of DMTS procedures often show a great deal of inter-participant variability and this can be a problem when interpreting the data. Adamson (1995) found variability across individual participant's data and variability across experimental groups has also been of concern (Huppert & Piercy, 1978; Kopelman, 1985; and Hart et al., 1987). As mentioned previously Huppert and Piercy (1978) compared forgetting in patients with Korsakoff's syndrome to that of an age-matched control group. They found that the performance of amnesic participants was typically lower than that of age-matched controls, for the shortest delay, when the experimental conditions were the same for both groups. They suggested that this made it difficult to compare the rate at which information was forgotten, across the groups, as the delay interval was increased. Huppert and Piercy (1978) attempted to equate initial performance and found that increasing the exposure time to stimuli resulted in similar levels of initial performance for a group with Korsakoff's dementia and an age-matched control group.

Kopelman (1985) and Hart et al. (1987) also attempted to equate initial levels of performance across experimental groups, in a manner similar to that of Huppert and Piercy (1978). Kopelman (1985) referred to this process as a titration. Kopelman (1985) and Hart et al. (1987) tested participants using a yes/no criterion test and if a participant was unable to obtain the required number of correct responses their exposure time to the stimuli was increased. This process was repeated until participants were able to obtain the required number of correct responses. The intention of these studies was to investigate whether the rate at which different experimental groups forgot information differed. Thus, they attempted to equate initial levels of performance by adjusting aspects of the experimental procedure so they could compare changes in performance as the delay interval was increased.

Wernham (1997), using a task similar to that used by Adamson (1995), developed a procedure which was designed to select the physical disparity

between sample and comparison stimuli, at which an individual participant reached a criterion level of discriminative performance. He also referred to this as a titration procedure. On each trial a single sample stimulus was presented in the centre of a computer screen, immediately after it disappeared from the screen two comparison stimuli were displayed. The participant was required to select the comparison stimulus that was the same size as the sample stimulus. This task was effectively a zero delay matching to sample task. At the beginning of the titration the difference between the sample and the non-matching comparison stimuli was small. After a pre-determined number of trials at that physical disparity, the percentage of correct responses was calculated by the computer. If the participant had not achieved the pre-set criterion level of accuracy the physical disparity was automatically increased. This process was repeated until the participant was performing at the pre-set criterion level of accuracy.

Wernham (1997) found that the titration procedure could be used to produce similar levels of accuracy, at zero delay, across participants. Wernham's (1997) titration procedure was designed using a single sample stimulus and was applied to a DMTS task where there was one sample stimulus. However, if the number of sample stimuli used effects accuracy on this task it is possible that different physical disparities might be reached if the titration used different numbers of sample stimuli. If this were so, then it might not be appropriate to use this single sample titration for a DMTS task where there is to be more than one sample stimulus.

Given these considerations the present experiment had two aims. The first was to confirm that using a titration procedure, similar to that developed by Wernham (1997) would result in similar levels of performance on a DMTS task, across participants. The second was to investigate whether the number of sample stimuli used for the titration task has an impact on the physical disparity reached for the pre-set criterion level of accuracy. A total of 100 trials were used here, as Wernham (1997) found that almost all participants in his study had reached criterion within 100 trials.

Method

Subjects

There were five participants in this experiment, all enrolled in a first year psychology course. The participants were recruited through a notice placed on a notice board in the Psychology Department. Each participant received a 1% course credit for each hour of participation, irrespective of their performance on the experimental task.

Apparatus

The experiment was conducted using Digital Venturis 100 computers. The stimuli were presented on a 35 cm SVGA monitor. The computer arranged the experimental events and recorded the participants' responses.

Stimuli

The stimuli were filled white disks 49 and 98 pixels in diameter, presented on a blue background. When there was one sample stimulus it was 98 pixels in diameter. The non-matching comparison stimuli differed in size from the sample stimuli by a proportion of the diameter of the sample stimulus. For each of the two sample stimuli there were two possible non-matching sample stimuli, one larger and one smaller than the sample stimulus. For example, consider a sample stimulus of 49 pixels in diameter and non-matching sample stimuli which differed by a proportion of 10% from this. The diameter of the larger non-matching sample stimulus was 54 pixels in diameter (rounded to the nearest pixel), and the diameter of the smaller non-matching sample stimulus was 44 pixels in diameter (rounded to the nearest pixel). The proportional difference to be used during the DMTS phase of the session was selected for each participant during the initial phase of the experiment and is reported in the results section.

Procedure

At the beginning of the experiment the participants were given an instruction sheet describing what was required of them to complete the experimental task. The instructions were as follows:

> In this experiment, you will be presented with a circle in the center of the computer screen. This circle will remain on the screen for a brief period of time, after which it will disappear. After a varying delay two circles will appear, one on the left of the screen and one on the right of the screen. One of these circles will be the same size as the circle you originally saw, the other one will be either slightly larger or smaller. You will be required to choose the circle that you think is the same size as the original circle. If you think the circle on the left is the same size press the left mouse key. If you think the circle on the right is the same size press the right mouse key. This process will be repeated until the experiment is completed. If you have any questions please ask the experimenter before the experiment begins.

After the participants had read the instructions, the experimenter asked whether there were any questions. If they had any questions, the experimenter answered them before the experiment began. Each participant was then seated at a computer with a screen presenting the following instructions:

> If you want to choose the circle on the left use the LEFT mouse button. If you want to choose the circle on the right use the RIGHT mouse button. Please enter your ID number:

After typing in their Student Identification Number they were instructed to press any key to initiate the first trial.

The experimental session began with a titration procedure which was designed to adjust the proportional difference between the non-matching comparison stimuli and the sample stimuli. This phase of the experiment used a single sample stimulus, 98 pixels in diameter. Each trial commenced with the sample disk presented in the center of the computer screen for 1.5 s. After a delay

of 0.5 s the choice phase started with the presentation of two comparison disks, one centred on the left half and the other centred on the right half of the screen. One of the comparison stimuli was the same size as the sample stimulus, while the other was randomly selected to be either larger or smaller than the sample stimulus. The side on which the matching and non-matching comparison stimuli appeared was also randomly selected. The two comparison stimuli remained on the screen until the participant pressed one of the two mouse buttons. No feedback, regarding whether performance was correct or incorrect, was provided. The next trial began immediately after a participant had pressed a mouse button. There were 100 trials available for this one-sample titration.

The proportional difference between the sample stimulus and the non-matching stimuli was initially set at 1% of the diameter of the sample stimulus. The participants initially completed 10 trials at the 1% proportional difference. If they achieved less than 100% correct on the 10 trials, the proportional difference was increased by 2%. This process was repeated until the participant was 100% correct on 10 trials. A further 20 trials were then completed at this proportional difference. If the participant was not 100% correct on these 20 trials the proportional difference was increased by 2% for 10 more trials and the process was repeated. If the participant was correct on all 20 trials before they had completed 100 trials they completed the remaining trials at that physical disparity. If this occurred and the participant made an error on any of these remaining trials the physical disparity was not increased. The one-sample titration finished after 100 trials and a two-sample titration then began.

The general procedure for the two-sample titration was the same as that used for the one-sample titration. It differed in that two sample stimuli, 49 and the 98 pixels in diameter, were used. The sample stimulus which appeared on each trial was randomly selected. Both sample stimuli were presented the same number of times. There were 100 trials available for the two-sample titration and participants completed the titration in the same way as the one-sample titration. At the completion of the two-sample titration the DMTS phase of the session began using the proportional difference of the two-sample titration. If the participant was not performing at 100% correct before they had completed the 100

trials the DMTS task began using the largest proportional difference reached in the 100 trials.

The general procedure for the DMTS task was the same as that used for the two-sample titration except for the inclusion of one of a set of delays (0.5, 2, 4, 8 or 16 s) between the presentation of the sample stimulus and the choice phase. There were a total of 160 trials during the DMTS procedure with 32 trials at each of the delay intervals. For each trial the delay interval was randomly selected without replacement so that each of the delay intervals occurred the same number of times. The sample and the non-matching comparison stimuli were randomly selected, without replacement, for each trial at each delay. During the comparison phase each sample stimulus was presented the same number of times with its larger and its smaller non-matching comparison stimuli. The experimental session ended when the participant had completed the 160 trials of the DMTS procedure. Once the experimental session had ended the participants were debriefed about the purpose of the experiment.

Results

Figure 1.1 shows the physical disparity between the sample stimulus and the non-matching comparison stimuli, as a proportion of the sample stimulus diameter, reached at the end of both the one-sample and the two-sample titrations. For all individuals the physical disparity was greater for the two-sample titration than for the one-sample titration. The proportional differences reached were 11, 13, 13, 13 and 15% for the one-sample titration and 13, 19, 19, 15, 19% for the two-sample titration for S1 to S5, respectively.

Figure 1.2 shows the correct or incorrect response made on each experimental trial plotted by the physical disparity on each trial, for each participant and for both titrations. Correct responses are indicated by a dot and incorrect responses are indicated by a plus and are shown for each of the physical disparities reached. In some cases the horizontal bar of the plus is obscured by the preceding and following correct response. S1 reached criterion for both the one and the two-sample titration. For the two-sample titration S1 was 100% correct

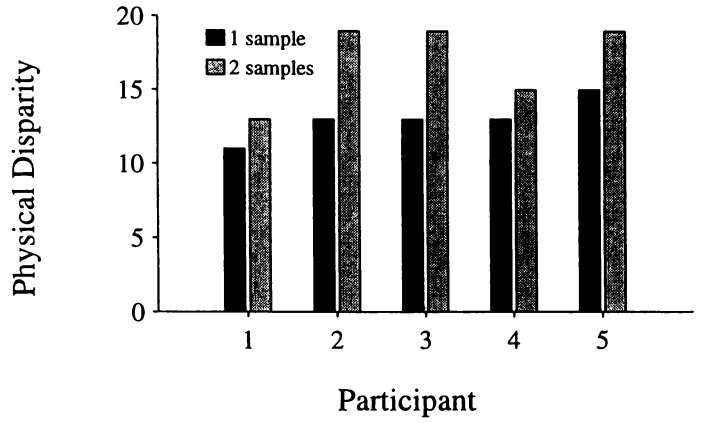

Figure 1.1. The physical disparity reached at the end of both the one- and the two-sample titrations for each of the participants.

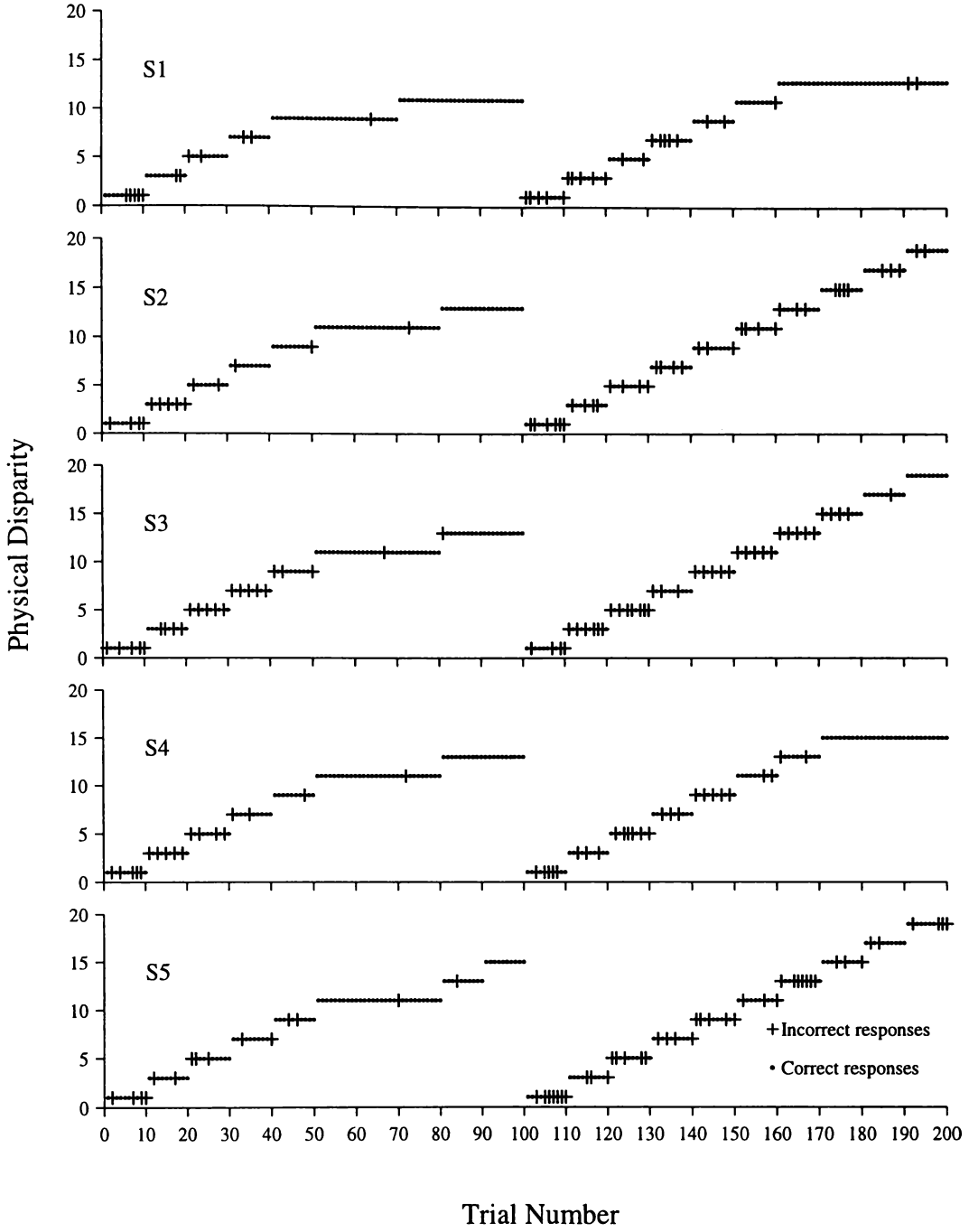

Figure 1.2. Correct (dot) and incorrect (plus) responses, for both titrations, as a function of experimental trial and physical disparity for each of the individual participants.

across the first 30 trials with the 13% disparity. Thus, S1 reached the pre-set level of accuracy with the 13% disparity. S1 continued with this physical disparity until the maximum of 100 trials had been completed and made two errors. This meant that over the last trials of this titration S1 did not perform at 100% correct. However, as S1 had already reached the criterion level of accuracy these errors did not result in an increase in the physical disparity. S4 reached criterion for the two-sample titration only. S2, S3 and S5 had not reached the criterion by the end of the 100 trials with either titration.

For the participants who did not reached the performance criterion, over the required number of trials, the percentage of correct responses was calculated for the final physical disparity. For the one-sample titration, S2 and S4 were 100% correct over 20 trials, S5 was 100% correct over 10 trials and S3 was 95% correct over 20 trials. For the two-sample titration, S3 was 100% correct, while S2 was 80% correct and S5 was 70% correct over 10 trials. For S3 the one-sample titration procedure did not function correctly for the physical disparity which they had reached at the end of the 100 trials. S3 was incorrect on the first of a series of ten trials, indicated by the plus at the beginning of the last set of trials. This meant that S3 was not performing at the criterion level of accuracy, across these 10 trials, and thus, the physical disparity should have increased but it did not.

Figure 1.3 shows percent correct as a function of delay interval for the DMTS task for each of the five participants and the average of these. For all individuals, accuracy was high. S5's accuracy was lower at 0.05 s than for any of the other participants. At a delay of 0.05 s, accuracy for S1 to S4 ranged from 91% to 97% correct and for S5 it was 73% correct. There were only small decreases in accuracy as the delay interval increased for most participants. The average data were similar to the individual data in that initial accuracy was high and accuracy decreased only slightly as the delay interval increased.

Figure 1.4 shows logit p, calculated using Equation 3, for each of the five participants and averaged across all five participants, as a function of delay. Logit p is a measure which is not bounded at the upper end of the scale as is percent correct. This means that, because it is not constrained, small differences

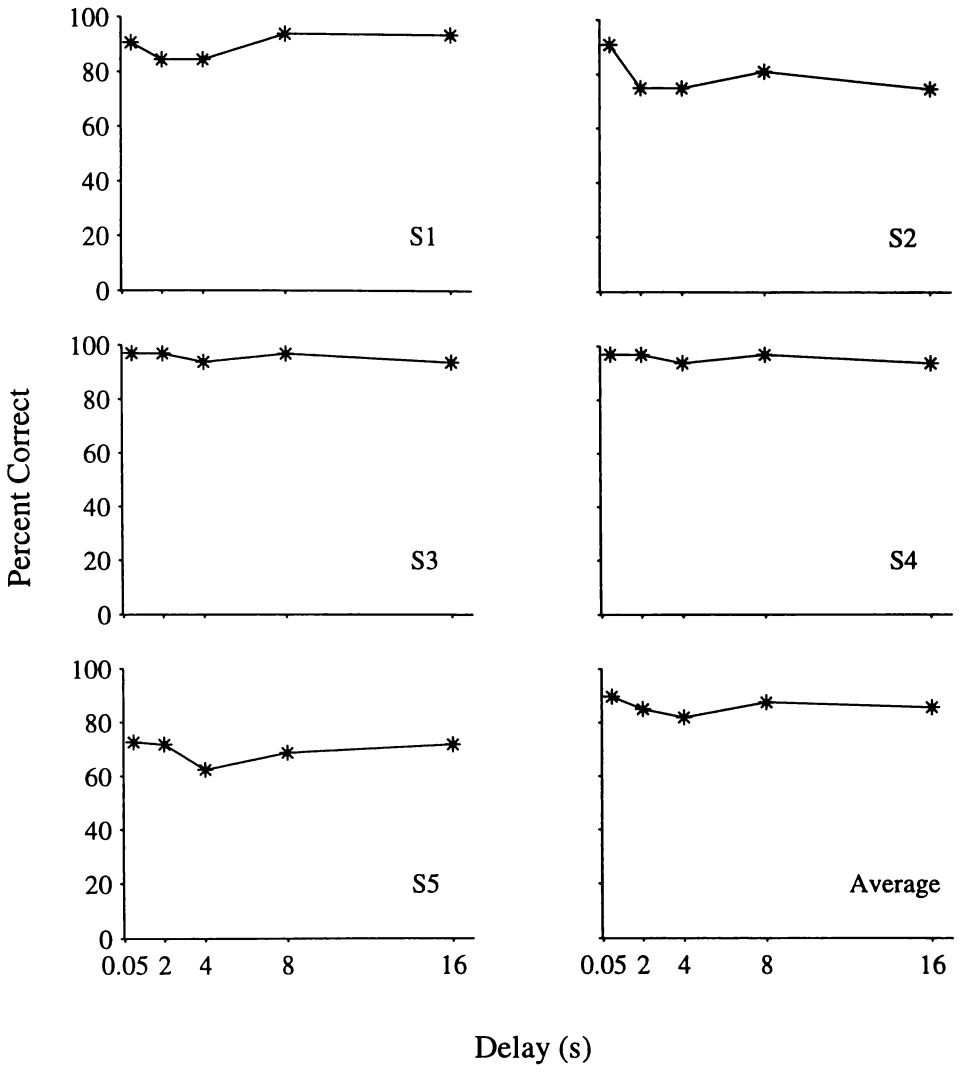

Figure 1.3. Percent correct for the each participant and the average of these, plotted as a function of delay.

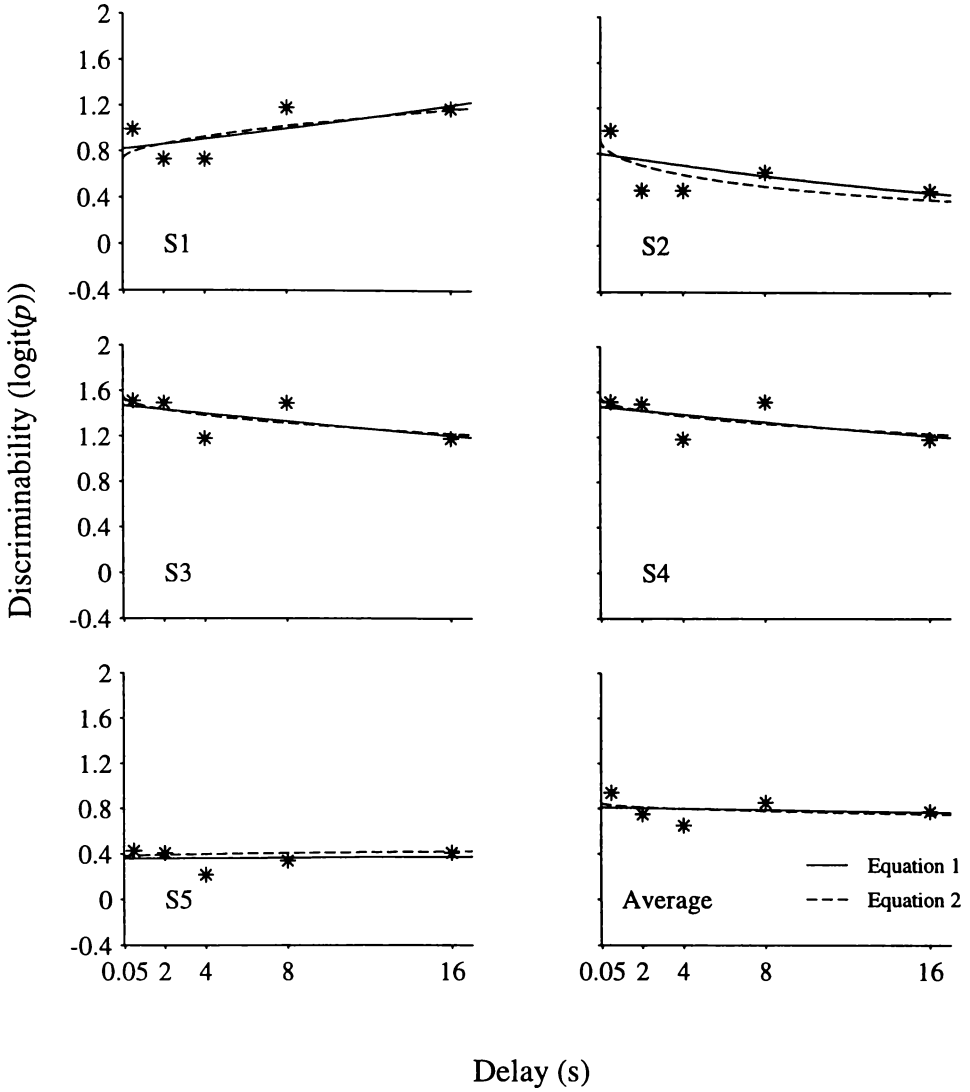

Figure 1.4. Estimates of discriminability (logit p) for the Individual participants and the average of these, plotted as a function of delay. The data are fitted by negative exponential functions (Equation 1 and Equation 2).

in accuracy, at the higher end of the scale, appear to be larger than they do when percent correct is used. The same trends are evident in the logit *p* data as were evident in the percent correct data.

Negative exponential functions (Equation 1, solid line and Equation 2, dashed line) were fitted to the logit *p* (Equation 3) data using non-linear regression and these functions are shown on Figure 1.4. Table 1.2 shows the values obtained, for both functions, for the parameters *a* and *b* for both the individual and the average data. In some cases it is difficult to see the dashed line used to fit the function for Equation 2 as it is very similar to the function provided by Equation 1 which is plotted as a solid line. For the individual data the *a* values, for both of the functions, show that initial accuracy was high and the *b* values show that there was very little decrement in accuracy with increasing delay. The lowest *a* was for S5 and this participant's accuracy was also the lowest at the end of the titration. The *b* values, for both functions, for S1 are negative as a result of the high accuracy at the 8 and 16 s delays. There was very little difference between the parameter values, *a* and *b*, given by the two negative exponential functions (Equations 1 and 2).

Table 1.2 also shows the standard errors of estimate and the percentages of variance accounted for (VAC) by both of the fitted functions. In all cases the percentage of VAC was low for both of the fitted functions. For all participants, except S2, there was very little difference in the percentage of VAC by the two fitted functions. The standard errors for both of the functions were small for all participants and there was little or no difference in these standard errors across the two functions.

The data for each of the participants were analysed for any changes in performance across the experimental session. To do this the data from the DMTS procedure were divided into two blocks each containing 80 trials. Each block was then analysed as a function of delay. Figure 1.5 shows that there was no systematic change in accuracy across the two blocks. S4 showed the greatest difference between the two blocks with accuracy on the second block of trials lower for all delays except for the 16 s delay.

Table 1.1 *Estimates of a, b, the standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions Equation 1(simple negative exponential) and Equation 2(negative exponential (sqrt(t))) for each of the individual participants and the average of these.*

	Equation 1				Equation 2			
	a	*b*	VAC%	Std Err	*a*	*b*	VAC%	Std Err
S1	0.82	-0.02	48	0.16	0.73	-0.12	42	0.17
S2	0.74	0.03	24	0.20	0.94	0.21	40	0.18
S3	1.47	0.01	34	0.14	1.56	0.06	35	0.14
S4	1.47	0.01	32	0.14	1.56	0.06	33	0.15
S5	0.36	0.00	0	0.08	0.38	0.03	2	0.08
Average	0.81	0.00	3	0.10	0.85	0.03	9	0.10

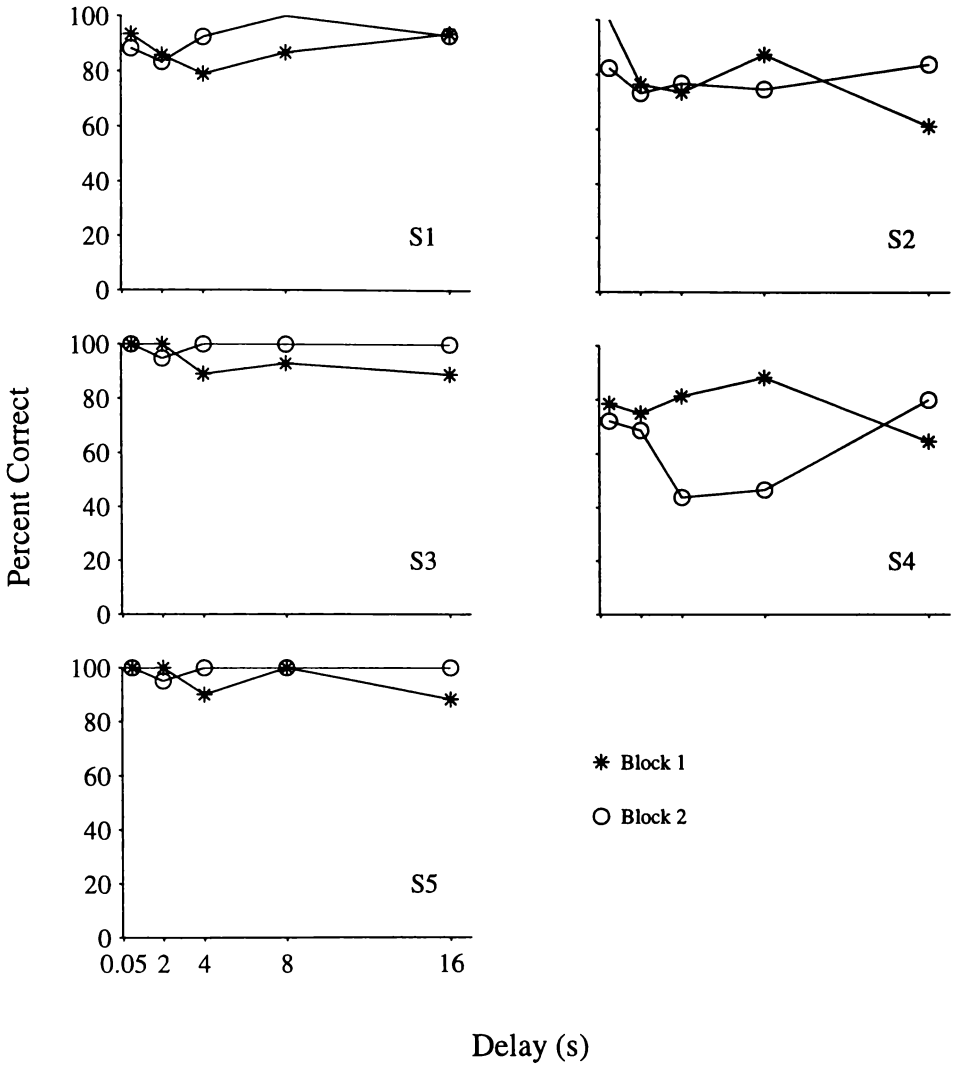

*Figure 1.5.*Percent correct on the first and second block of 80 trials for the individual participants as a function of delay (see text for details).

Discussion

Adamson (1995) used a DMTS task to study human memory and found that inter-subject variability in *a* made it difficult to interpret the data. The present study used a DMTS task similar to that of Adamson (1995) with the addition of a titration, which was intended to reduce inter-subject variability in *a*. In the present experiment initial discriminability was generally similar for all of the participants. Thus, there was less inter-subject variability in *a* than might have been expected for a DMTS task if a titration had not been used, given the results of Adamson (1995). This supports Wernham's (1997) finding that using a titration to adjust the physical disparity of the sample stimuli results in similar levels of initial discriminability, *a*, across participants. Additionally, here the physical disparity reached at the end of the titration was larger, for all of the participants, when two sample stimuli were used than when one sample stimulus was used.

The purpose of the titration was to allow participants to enter the DMTS task with the physical disparity which gave a pre-set level of accuracy at the shortest delay. In this respect the titration was not entirely successful here as only S1 and S4 were performing at the criterion level at the end of the two-sample titration. Most participants were not performing at the criterion level at the end of either titration. Despite this, accuracy at the end of the titrations with the exception of S5, was generally high and similar across participants. Thus, the criterion might have been reached had more trials been allowed in the titration.

One change which could be made to the present titration to allow participants to reach the pre-set criterion is to increase the initial physical disparity between the sample and the non-matching comparison stimuli. This should reduce the number of trials required for participants to reach the pre-set criterion level of accuracy. In the current procedure the initial physical disparity was 1% and the smallest physical disparity reached for either titration was 11%. This suggests that there is room to increase initial physical disparity. The initial physical disparity could be greater than 1% but should be smaller than 11% of the diameter of the sample stimulus.

A second possible change would be to have more than 100 trials allowed in the titration procedure. The number used here was based on Wernham's (1997) finding that most participants reached a stable level of performance within 100 trials. However, this was not so in the present study.

In the present experiment the shortest delay used for the DMTS task (0.5s) was longer than the delay used for the titration (0.05s). Given that accuracy typically decreases as the delay is increased it might be expected that the use of this slightly longer delay, for the DMTS task, would result in a lower level of accuracy than that attained for the titration. This appears to be what happened here, with the level of accuracy attained at the shortest delay for the DMTS task being lower than the level of accuracy attained at the end of the two-sample titration, for three out of the four participants. However, this was not so for the remaining two participants. Thus, the differences between the shortest delay interval used in the DMTS task and the delay used in the titration does not account for these differences in accuracy. However, given that the different delay intervals may have contributed to these differences it might be better to have the shortest delay for the DMTS task the same as the delay used for the titration.

In the present experiment increasing the number of sample stimuli resulted in an increase in the physical disparity reached at the end of the titration for all of the participants, despite the fact that in most cases the criterion was not reached for either titration. Thus, with the same physical disparity one might expect the initial discriminability to be lower with a two sample DMTS task than with a one sample DMTS task. This provides support for Adamson's (1995) suggestion that increasing the number of sample stimuli results in a decrease in initial discriminability. Thus, if a one-sample titration was used to select the physical disparity for a two sample DMTS procedure you might expect initial discriminability on the DMTS task to be lower than the pre-set criterion level of accuracy for the titration. However, it is not possible to say this with absolute certainty as three of the five participants did not reach the criterion level of 100% for the present experiment.

White (1985) found that increasing the physical disparity between two sample stimuli resulted in an increase in initial discriminability, *a* but did not

change *b*. The results of both of the titrations showed that as the physical disparity between a sample stimulus and its respective non-matching comparison stimuli was increased matching accuracy also increased. Thus, the present result is consistent with White (1985).

White Ruske and Colombo (1996) suggested that Equation 2 (negative exponential (sqrt(*t*))) should be used to quantify changes in accuracy as the delay is increased as it provides a better fit than Equation 1 (simple negative exponential). Thus, for the present experiment, both the simple negative exponential (Equation 1) and the negative exponential (sqrt(*t*)) (Equation 2) were fitted to the data. For both functions it was generally true that the percentage of VAC was lowest when the *b* values were small and was largest when the *b* values were large. This is because when *b* is small there is little variance in the data to be accounted for by the function. In the case of S2, where *b* was large and there was variance to account for, the percentage of VAC for by the negative exponential (sqrt(t)) was almost double that of the simple negative exponential. For the remaining participants the fits provided by the simple negative exponential and the negative exponential (sqrt(*t*)) do not differ markedly.

In conclusion the present experiment suggests that using a titration procedure can result in similar patterns of performance on a DMTS task, across individual participants. The present finding also suggests that using two sample stimuli rather than one sample stimulus in a titration procedure results in an increase in the physical disparity between a sample stimulus and its non-matching comparison stimuli required to reach a criterion level of accuracy.

Experiment 2

The previous experiment showed that although the participants were performing at different levels of accuracy at the end of the titration, there was less inter-subject variability, in *a*, on the DMTS task than might have been expected if a titration had not been used. It was suggested that some of the remaining differences in initial discriminability may have occurred because not all participants had reached 100% correct by the end of the titration. These differences in accuracy may be reduced if the titration procedure was modified to allow all of the participants to reach the criterion.

In the previous experiment it was suggested that the physical disparity should be increased above 1% but be smaller than 11%. Therefore, in this next experiment the initial physical disparity was increased to 4% of the diameter of the sample stimulus. The titration for the previous experiment ended after 100 trials had been completed and most participants had not reached the pre-set criterion level of accuracy. The maximum number of trials used during the present titration procedure was increased from 100 to 200 trials. It was hoped that doubling the number of trials would allow participants to reach a physical disparity at which they could perform at criterion. It was also decided that the titration would be modified so that if participants were performing at criterion before 200 trials had been completed the titration ended and the DMTS task began immediately. The maximum physical disparity that could be reached at the end of the titration was set to be 28% as when the physical disparity between the sample and the non-matching comparison stimuli was larger than this both stimuli may not fit on the screen during the comparison phase.

The present experiment investigated further the impact that the number of sample stimuli used in a titration procedure has on performance on a DMTS task. This was done by using two titrations as in Experiment 1, one with 1 sample stimulus and one with 16 sample stimuli. The disparity reached at the end of each of the titrations was then used in a DMTS task. Three different sample-set sizes 16, 4 and 1, were used in the DMTS task. Each of the sample sets was used in a DMTS task, once with the physical disparity reached at the end of the 1-sample

titration and once with that reached at the end of the 16-sample titration. Thus, it was possible to compare performance on these DMTS tasks with two different physical disparities (provided the two titrations gave different results).

Method

Participants

Five first year psychology students participated in this experiment. Each participant received course credit towards a first year psychology course, irrespective of their performance on the experimental task and irrespective of whether or not they completed the experiment.

Apparatus

The apparatus used was the same as that used in Experiment 1.

Stimuli

Filled white disks of varying sizes, presented on a blue background, were used as stimuli. The sample stimuli were drawn from a pool of 16 disks ranging in size from 60 pixels to 195 pixels in diameter, in steps of 9 pixels. There were three different sized sample sets with 1, 4, or 16 stimuli. When there was one sample stimulus it was 123 pixels in diameter. The 4 sample-stimulus set contained stimuli 114, 123, 132 and 141 pixels in diameter. The 16 sample stimulus set contained all of the stimuli between and inclusive of 60 to 195 pixels in diameter, in steps of 9 pixels. The non-matching comparison stimuli differed in size from the sample stimuli by a proportion of the diameter of the sample stimulus. For each sample stimulus there were two possible non-matching sample stimuli, one larger and one smaller than the sample stimulus.

Procedure

The general procedure used was the same as that used for Experiment 1. The initial part of the session was a one-sample titration procedure using a 0.05 s delay. There were 200 trials allowed in the titration and the maximum physical disparity that a participant could reach was 28%, of the diameter of the sample

stimulus. When a participant was 100% correct on 10 trials and was also 100% correct on the next 20 trials the titration procedure ended. Thus if the participant reached the pre-set criterion of 100% over a total of 30 trials the titration procedure ended regardless of whether the participant had completed 200 trials or not. The titration procedure also ended if the participant had completed the 200 trials before getting 100% correct over a total of 30 trials. If a participant reached the maximum physical disparity of 28% the titration continued, with this physical disparity, until the 200 trials were completed.

The one-sample titration was followed immediately by a DMTS task where the delays between the presentation of the sample stimulus and the choice phase were 0.05, 4, and 8 s and three sample set sizes were used. The physical disparity that had been reached at the end of the one-sample titration by each of the participants was used as the disparity between the sample stimuli and the non-matching comparison stimuli.

The delay intervals were blocked and appeared in the following order, 0.5s, 4s and 8 s with a total of 96 trials at each delay interval. The sample sets appeared in the following order, 16, 4 and 1, for each of the delay intervals in blocks of 32 trials. When each of the three sample-set sizes had been completed at a particular delay the next delay started. There were a total of 96 trials for each of the sample-set sizes across all of the delays. The number of times that each sample stimulus was presented at each of the delays varied for each sample-set size. For the 16-sample stimuli each sample was presented twice, for the 4-sample stimuli each sample was presented 8 times and for the 1-sample stimulus the sample was presented 32 times. The sample stimulus used on each trial was randomly selected for the 16- and the 4- sample sets. Each sample stimulus had a larger and a smaller non-matching comparison stimulus each of which was randomly presented an even number of times, on trials with that sample stimulus. The DMTS phase ended when the participant had completed 288 trials.

At the completion of this DMTS procedure a 16-sample titration procedure began using a similar procedure to the 1-sample titration. It differed from the 1-sample titration in that 16 stimuli 60 to 195 pixels in diameter were used as sample stimuli. Each of the 16 sample stimuli was presented as the sample

stimulus twice for each of the delays. At the completion of the 16-sample titration procedure a second DMTS procedure began.

The procedure used for the second DMTS task was the same as that used in the first DMTS task except the physical disparity that had been reached at the end of the 16-sample titration was used for each of the participants. The physical disparity between the sample and the non-matching comparison stimuli was the same across all three sample-set sizes for each participant. Once the experimental session had ended the participants were given details about the purpose of the experiment.

Results

Figure 2.1 shows the physical disparity reached at the end of the 1- and the 16-sample titrations. At the end of the one-sample titration these physical disparities were 14%, 28%, 12%, 24% and 12% for S1 to S5, respectively. At the end of the 16-sample titration the physical disparities reached were larger, with the exception of S2, than those reached at the end of the 1-sample titration. For the 16-sample titration all of the participants reached the maximum physical disparity of 28%.

Figure 2.2 shows correct and incorrect responses as a function of trial number and physical disparity for both the 1- and the 16-sample titrations, as in Experiment 1 (Figure 1.2). All of the participants reached the 100% criterion for the 1-sample titration procedure. For the 16-sample titration, all of the participants reached the maximum physical disparity of 28% but none of them reached the criterion level of accuracy for this physical disparity. The level of accuracy at which the participants were performing at the end of the 16-sample titration was calculated over all trials on which the physical disparity was 28%. S1 to S5 completed 50, 90, 90, 70 and 50 trials, respectively, at the physical disparity of 28% and obtained 86, 64, 78, 70 and 88% correct, respectively over these trials.

Figure 2.3a shows percent correct, for each of the five participants, and Figure 2.3b shows the average of these, as a function of delay interval and

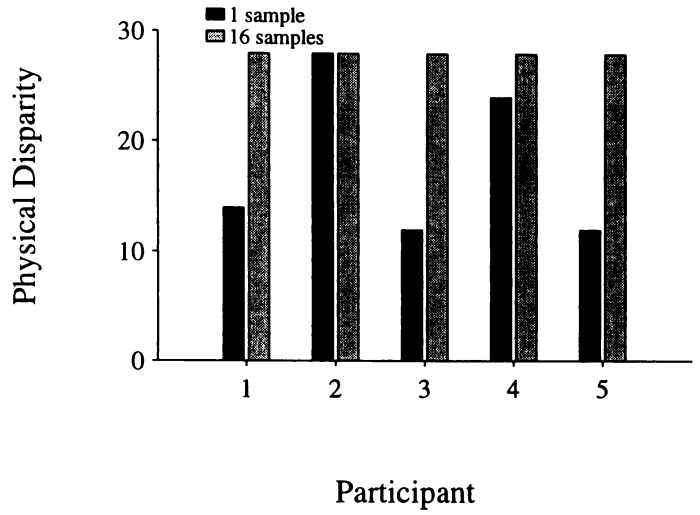

Figure 2.1. The physical disparity reached at the end of both the 1- and the 16-sample titrations for each of the participants.
of the participants.

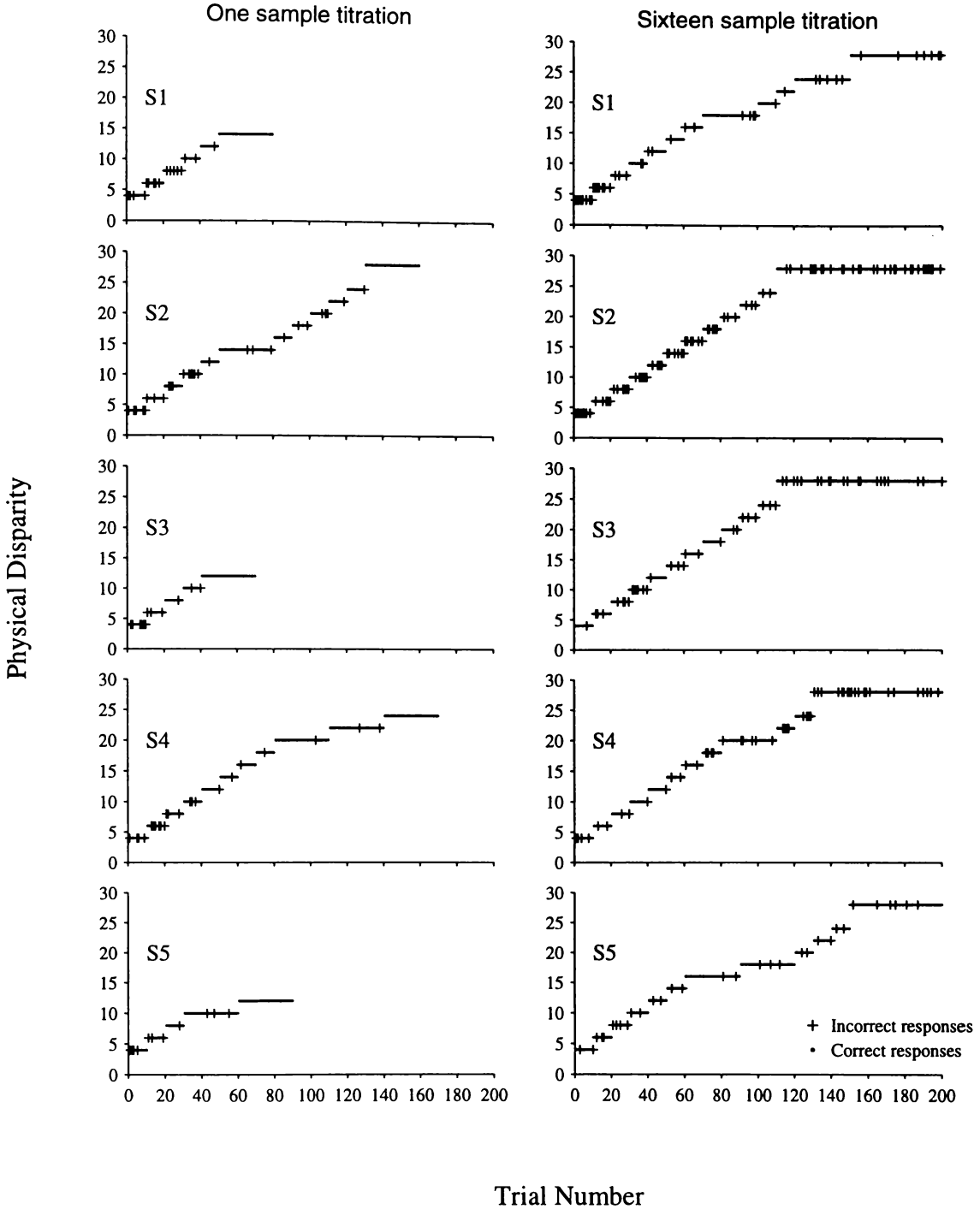

Figure 2.2. Correct (dot) and incorrect (plus) responses, for both titrations, as a plotted as a function of both experimental trial and physical disparity for each of the participants.

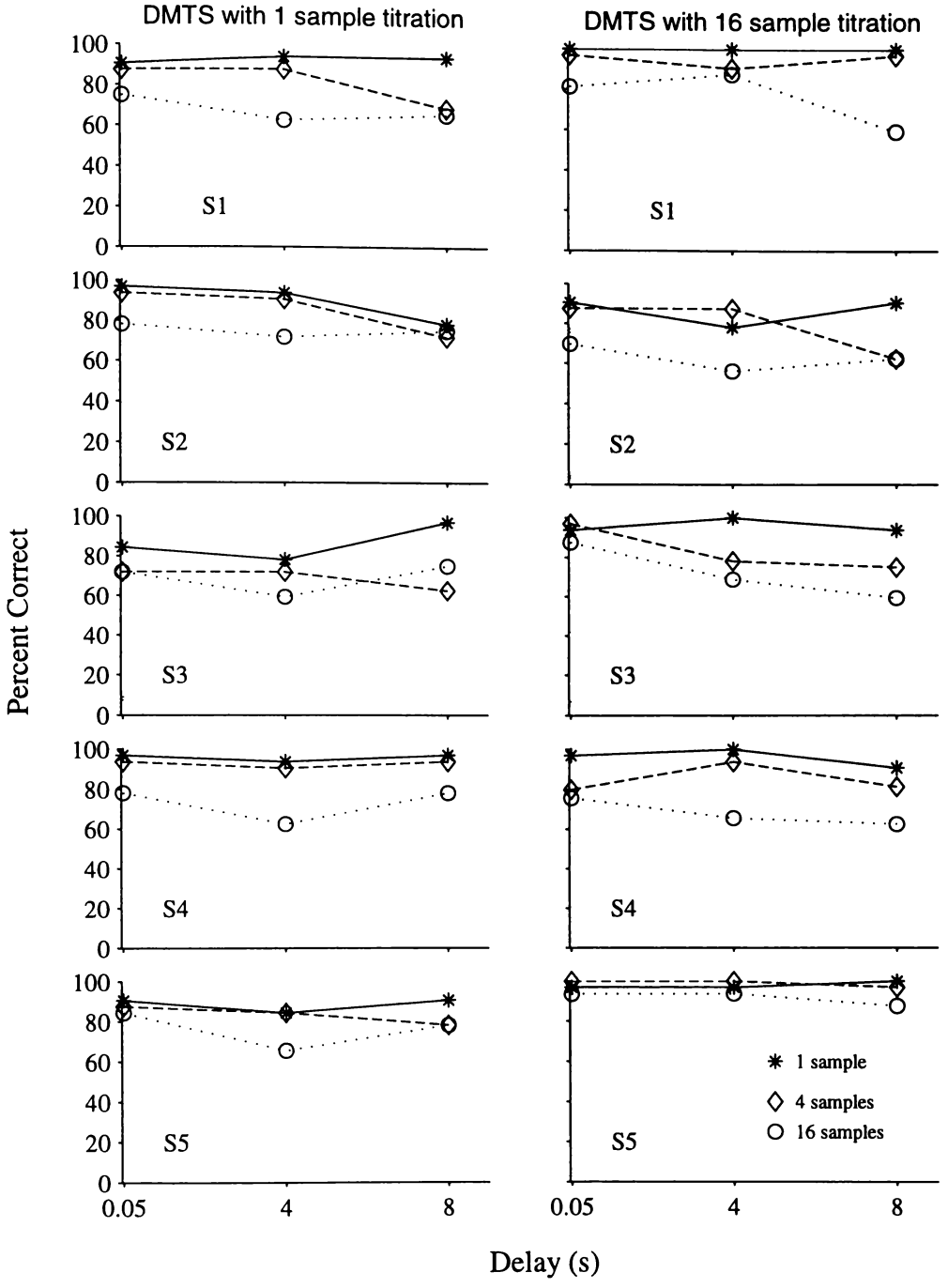

Figure 2.3a. Percent correct plotted as a function of delay and sample-set size for the indvidual participants.

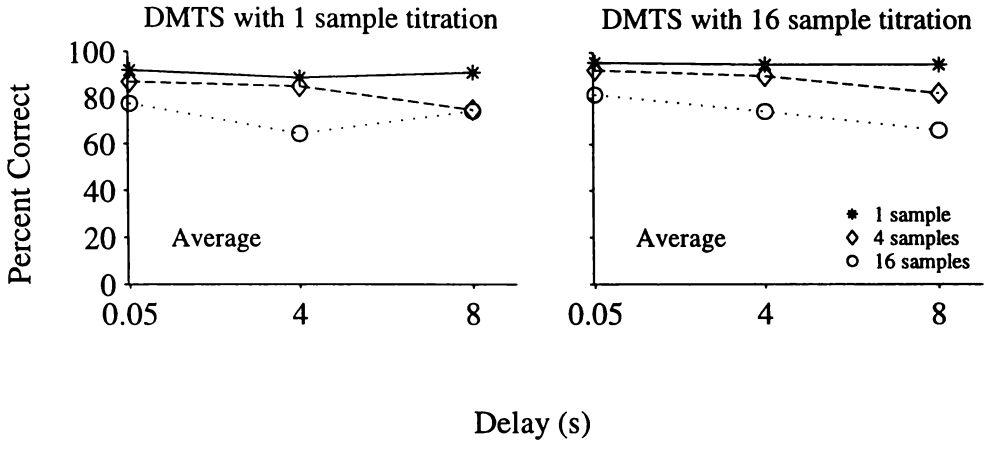

Figure 2.3b. Percent correct as a function of delay and sample-set size for the average data.

sample-set size for each of the two DMTS tasks. For all individuals accuracy decreased slightly as the delay interval was increased for all of the sample-set sizes and for both of the DMTS phases. Figure 2.3a also shows that accuracy, for both of the DMTS tasks, tended to decrease as the sample-set size was increased. For S1 and S4 percent correct decreased systematically as the number of sample stimuli was increased for all delay intervals and for both of the DMTS tasks. For the remaining participants percent correct tended to be similar for the 1 and the 4 sample-set sizes and was lowest for the 16 sample-set size for both of the DMTS phases, with the exception of some delay intervals. Percent correct, for each of the sample-set sizes, was typically slightly higher when the 16-sample titration had been used to select the physical disparity, than when the 1-sample titration had been used.

The average data are similar to the individual data in that accuracy decreased systematically as the number of sample stimuli was increased for both DMTS tasks, with the exception of the 16 s delay for the first DMTS task. Percent correct was slightly higher for each of the sample-set sizes when the physical disparity had been selected using the 16-sample titration.

Figure 2.4a shows logit *p*, calculated using Equation 3, for each of the five participants and Figure 2.4b shows this for the averaged data as a function of sample-set size and delay interval for both of the DMTS tasks. For the second DMTS task four of the participants, the exception being S2, obtained 100% correct for at least one delay interval for the one sample condition and S5 also obtained 100% correct during the four sample condition. This means that logit *p* was undefined at these points and it was not possible to calculate it without using some form of mathematical correction. The Hautus (1995) correction was applied, in which 0.5 was added to both the number of correct responses and the number of error responses for all of the individual and the averaged data. As in Experiment 1, using logit *p* allows small differences at the high end of the percent correct scale to be seen more clearly.

For both DMTS tasks, logit *p* tended to decrease as the number of sample stimuli was increased, at most delay intervals. Logit *p* tended to be higher for each of the sample-set sizes when the physical disparity had been selected using

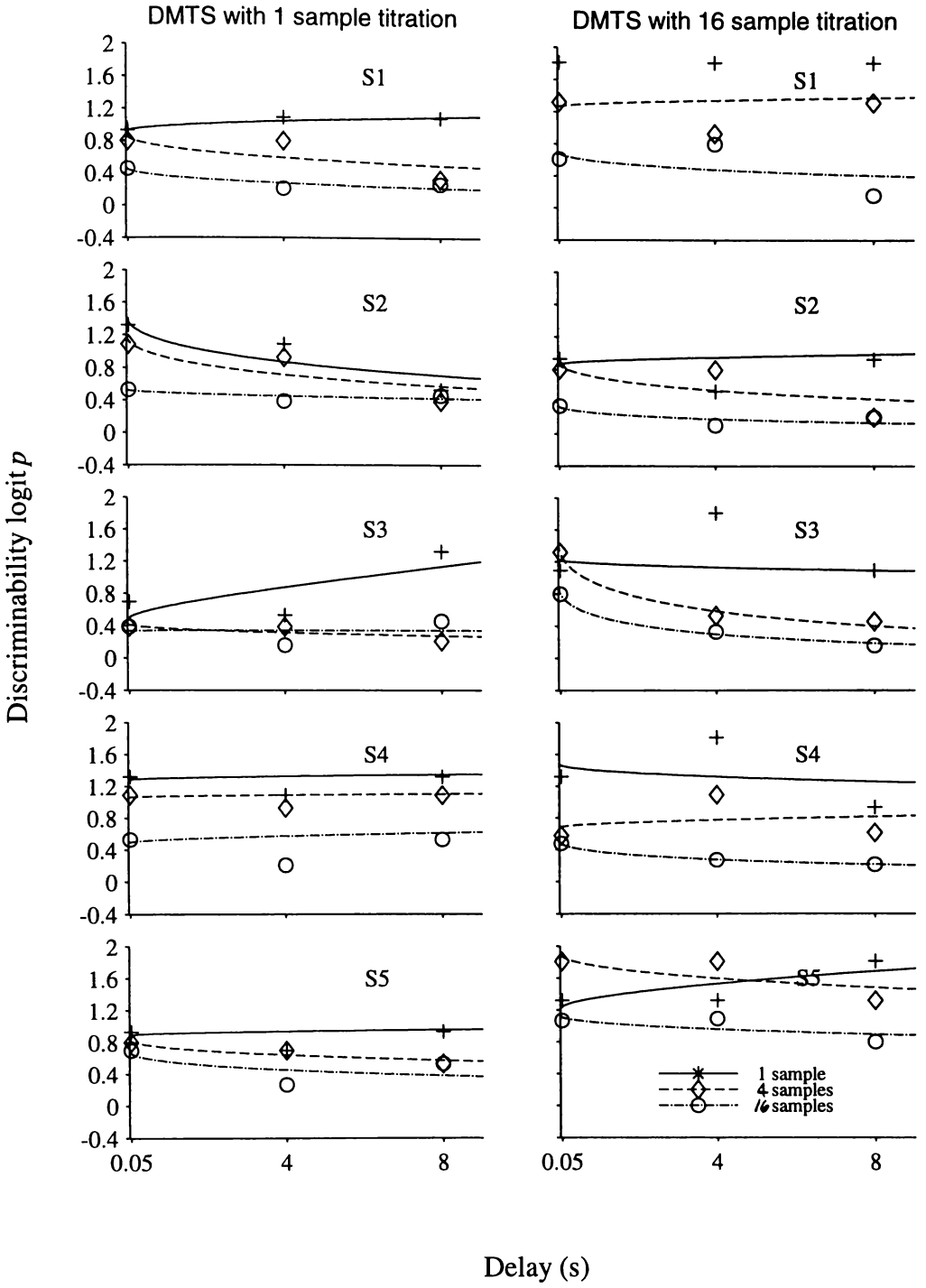

Figure 2.4a. Logit *p* plotted as a function of delay and sample-set size size for the indvidual participants. The fitted functions are negative exponentials (Equation 2).

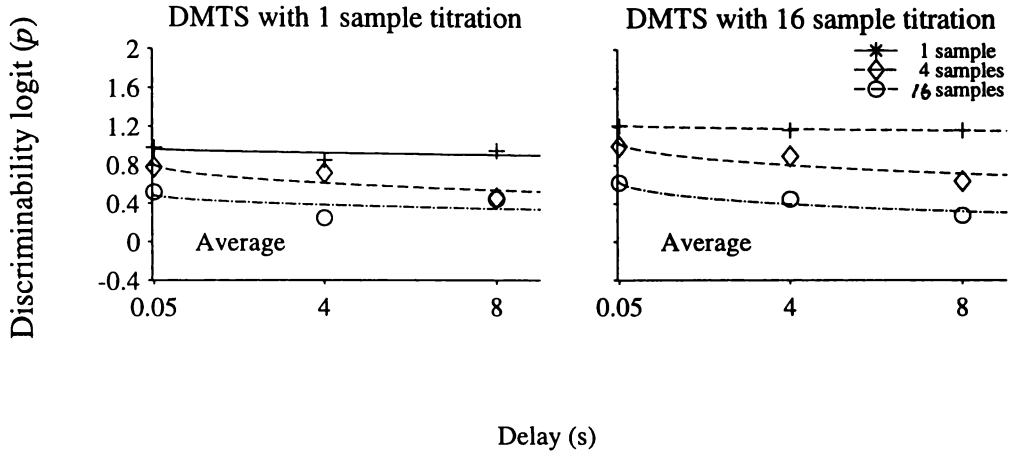

Figure 2.4b. Logit *p* plotted as a function of delay and sample-set size size for the averaged data. The fitted functions are negative exponentials (Equation 2).

the 16-sample titration (second DMTS task) than when it had been selected using a 1-sample titration. The averaged data were generally representative of the individual data in that logit *p* tended to decrease as the number of sample stimuli was increased. Accuracy, for the averaged data, also tended to be higher for each sample-set size for the second DMTS task.

A repeated measures ANOVA was carried out on these data for each of the DMTS phases of the experiment. Prior to the ANOVA Mauchly's test of sphericity was used to test for homogeneity of variance. For neither DMTS phase was this significant for delay (Phase 1 W = 0.34, Phase 2 W = 0.44), sample number (Phase 1 W = 0.89, Phase 2 W = 0.70) or the interaction (Phase 1 W = 0.02, Phase 2 W = 0.57) ($\alpha > 0.05$). For the first DMTS phase the main effect of delay (F = 2.27, df = 2.00, 8.00) was not significant but the main effect of sample number (F = 29.79, df = 2.00, 8.00) and the interaction (F = 5.27, df = 4.00, 16.00) were statistically significant ($\alpha < 0.05$). For the second DMTS phase the main effect of delay (F = 9.61, df = 2.00, 8.00) and the main effect of sample number (F = 18.98, df = 2.00, 8.00) was significant but the interaction (F = 0.90, df = 4.00, 16.00) was not significant ($\alpha > 0.05$).

Negative exponential functions (Equation 1 and Equation 2) were fitted to the individual and the average logit *p* data using non-linear regression. Table 2.1 shows the values obtained, for both equations, for the parameters *a* and *b* for each the individual participants and for the averaged data. Table 2.1 also shows the standard errors of estimate and the percentages of variance accounted for (VAC) by the both fitted functions. There was generally little or no difference in the standard error of estimate and the percentage of VAC by the two functions and thus only the functions for the modified negative exponential are shown on the graphs Figure 2.4a and Figure 2.4b. For S1 a line could not be fitted to the data from the second DMTS task for the one sample-set size as there was no variance in the values obtained with increasing delay.

Accuracy, for both DMTS tasks, typically decreased slightly as the delay interval was increased as shown by the *b* values in Table 2.1. For the first DMTS task *a* decreased systematically as the number of sample stimuli was increased, for all of the participants. For the second DMTS task *a* decreased systematically for

Table 2.1 *Shows a, b, the standard error of estimate and the percentage of VAC by the functions Equation 1 and Equation 2, for both the individual and the averaged data.*

Equation 1		Condition one				Condition two			
	Sample N	*a*	*b*	Std Err	VAC %	*a*	*b*	Std Err	VAC %
S1	1	0.96	-0.02	0.04	73	*	*	*	*
	4	0.87	0.08	0.13	67	1.19	0.00	0.19	0
	16	0.43	0.09	0.06	63	0.72	0.09	0.21	37
S2	1	1.38	0.10	0.10	89	0.79	-0.00	0.19	0
	4	1.15	0.10	0.12	84	0.88	0.10	0.17	64
	16	0.50	0.02	0.05	29	0.32	0.10	0.08	40
S3	1	0.50	-0.11	0.19	68	1.33	0.00	0.34	0
	4	0.42	0.06	0.05	69	1.30	0.17	0.11	93
	16	0.29	-0.03	0.12	6	0.81	0.21	0.01	100
S4	1	1.24	0.00	0.11	0	1.53	0.03	0.33	17
	4	1.03	0.00	0.08	0	0.75	0.00	0.23	0
	16	0.43	0.00	0.15	0	0.47	0.11	0.02	96
S5	1	0.85	0.00	0.11	0	1.23	-0.04	0.10	79
	4	0.82	0.05	0.02	96	1.89	0.03	0.14	72
	16	0.61	0.05	0.16	19	1.12	0.03	0.07	69
Average	1	0.94	0.00	0.06	4	1.21	0.00	0.01	75
	4	0.81	0.06	0.05	84	1.03	0.05	0.04	92
	16	0.45	0.03	0.11	8	0.63	0.09	0.01	99

Equation 2

S1	1	0.92	-0.06	0.03	89	*	*	*	*
	4	0.89	0.20	0.16	49	1.26	0.03	0.18	5
	16	0.48	0.27	0.05	80	0.71	0.19	0.24	21
S2	1	1.44	0.25	0.16	76	0.86	0.05	0.18	5
	4	1.20	0.26	0.16	70	0.91	0.26	0.20	48
	16	0.53	0.08	0.04	49	0.36	0.33	0.06	60
S3	1	0.47	-0.31	0.25	45	1.23	-0.04	0.34	4
	4	0.42	0.15	0.06	51	1.45	0.45	0.05	99
	16	0.34	-0.00	0.13	0	0.91	0.55	0.02	99
S4	1	1.28	0.02	0.11	4	1.48	0.06	0.35	6
	4	1.06	0.01	0.15	4	0.68	-0.06	0.23	6
	16	0.49	0.08	0.15	5	0.51	0.31	0.00	99
S5	1	0.89	0.03	0.10	5	1.21	-0.01	0.14	59
	4	0.84	0.13	0.04	86	1.91	0.09	0.16	53
	16	0.69	0.21	0.14	38	1.13	0.08	0.09	49
Average	1	0.97	0.02	0.05	17	1.22	0.02	0.01	91
	4	0.83	0.15	0.08	69	1.06	0.14	0.07	78
	16	0.50	0.13	0.10	23	0.66	0.26	0.04	93

*Indicates that the negative exponential function could not be fitted to the data for this participant for this sample-set sizes.

S1and S4 as the number of sample stimuli was increased and for S2, S3 and S5 it decreased systematically, with the exception of the four sample set size. There were no systematic changes in *b*, for any participant, as the sample-set size was increased for either of the DMTS phases. There was a systematic decrease in *a* as the sample-set size was increased for the average data for both of the DMTS phases. For the averaged data there were only slight decreases in accuracy for both of the DMTS phases, as shown by the generally small *b* values in Table 2.1. A unidirectional Ferguson's test for trend (Ferguson, 1965) was carried out on the *a* parameters for each of the DMTS phases of the experiment. Both the first and the second DMTS phases gave a statistically significant effect for the number of sample stimuli ($\alpha < 0.05$). A unidirectional Ferguson's test for trend (Ferguson, 1965) was also carried out on the *b* parameters for each of the DMTS phases. It was found that *b* did not trend significantly as the number of sample stimuli was increased for either of the two DMTS phases ($\alpha < 0.05$).

The data for each of the participants was analysed for any changes in performance across the experimental session. Each participant's data, for each of the delay intervals, was divided into two blocks of 16 trials for each of the sample-set sizes and for each of the two DMTS phases. The data were then examined for any systematic change in accuracy, across the two blocks of 16 trials, as a function of delay interval and sample-set size. These are not presented as there were no systematic changes in accuracy across the two blocks for any of the participants.

Discussion

The results of the present study suggested that increasing the number of sample stimuli used in the titration resulted in an increase in the physical disparity required to reach the preset criteria for four out of the five participants. This result was consistent with the finding in Experiment 1 for one and two sample stimuli. As in the previous experiment, not all of the participants attained the pre-set criterion level of accuracy by the end of the titration. In the present experiment all attained the criterion level in the 1-sample titration but none did so

in the 16-sample titration. Thus, the changes made to the titration, here, did help with the one-sample titration. However, further changes would be necessary to allow participants to reach the pre-set criterion level of accuracy when large sample-set sizes, such as 16 samples, are used in the titration.

In this study if the physical disparity had reached 28% but the participants had not attained the criterion performance the titration continued, with this physical disparity, until performance either reached the criterion or 200 trials were completed. All of the participants reached 28% for the 16-sample titration, here, but none of the participants attained the criterion level of accuracy. This suggests that the maximum physical disparity, used for the titration here was not large enough when there were 16 samples. The total number of trials that each participant completed with a physical disparity of 28% ranged from 50 to 90 trials. Thus, the present result also showed that if the participant's had not attained criterion performance after the first 30 trials, with 28%, they did not do so across subsequent trials with the same physical disparity.

The titration was designed to produce a high level of accuracy, in this case 100% correct, at the shortest delay of a DMTS task. If the sample-set size used for the titration was also used for the DMTS task you might, then, expect that participants would achieve 100% correct for this sample-set size at the shortest delay of 0.5 s. Thus, for the first DMTS phase, here participants might have been expected to achieve 100% correct with 1 but not with the 4 or 16 samples. For the second DMTS task, here participants would not have been expected to achieve 100% correct for 16 samples. However, given the larger physical disparity selected with the 16-sample titration participants might be expected to achieve 100% correct with 1 and 4 samples. This was a potential problem with this procedure. If a participant was 100% correct then it would not be possible to calculate the correct to error ratio for this condition without some form of correction. Given this problem, it may have been better to titrate the sample-comparison difference to an accuracy criterion less than 100% correct.

However, for the first DMTS task, where the physical disparity was generated using the one-sample titration, none of the participants achieved 100% correct for any of the sample-set sizes at any of the delays. Thus, for the present

experiment accuracy on the first DMTS task tended to be lower than the 100% correct criterion, used in the titration, which was achieved by all participants. This finding was consistent with the results of Experiment 1, where none of the participants were 100% correct for any of the delay intervals. It was suggested there that this may have been because participants did not generally reach the criterion for the two-sample titration used there. It was argued that this made it unlikely that they would achieve 100% correct at any of the delays for the DMTS task. However, for the present experiment all of the participants had reached criterion for the one-sample titration and so this suggestion does not account for the present result.

On the second DMTS task most of the participants obtained 100% accuracy for at least one of the delay intervals with one sample but none obtained 100% accuracy with either 4 or 16 samples. Thus, accuracy with one sample was generally higher for the second DMTS task than it had been at the end of the 16-sample titration for which none of the participants had reached the criterion, of 100% correct. However, the larger physical disparity reached in the 16-sample titration may have resulted in participants obtaining 100% with the one-sample set, when they did not obtain 100% with the smaller physical disparity from the 1-sample titration. If the criterion level of accuracy had been achieved for the 16-sample titration, then performance, on the DMTS task may have been at or near 100% correct, for all delays with the 1 and the 4 sample sets. Thus, if the number of sample stimuli are to be varied in a DMTS task and the intention is to examine the effects of sample number on DMTS performance, the titration procedure should probably use the smallest sample-set size. If the intention was for accuracy for all sample-set sizes to be similar for the shortest delay of a DMTS task, then it would be necessary to have a titration for each of the sample set sizes used for each DMTS task.

As mentioned earlier, logit p is undefined where a participant is 100% correct and some form of mathematical correction must used to calculate logit p for these points. Here the mathematical correction used was the Hautus (1995) correction. As suggested by Snodgrass and Corwin (1988), in order for logit p to be consistent across participants, the Hautus (1995) correction was applied to the

data of all participants, irrespective of whether or not they were not obtained 100% correct. Given that this procedure changes the logit *p* value in a systematic fashion the Hautus (1995) correction will be applied to all data in the remaining experiments here, to allow comparisons across this and future experiments.

The present study showed that accuracy tended to decrease systematically as the number of sample stimuli was increased in the DMTS task. This finding was reflected both in the percent correct analysis and in the *a* values obtained for the fitted functions. Worsham (1975) and Miskin and Delacour (1975) found matching accuracy increased as the number of sample stimuli was increased. The present result did not provide support for this. Nor did this result provide support for Etkin and D'Amato (1969) and Mason and Wilson's (1974) finding that increasing the sample-set size had no effect on matching accuracy.

Adamson (1995) found that increasing the number of sample stimuli resulted in a decrease in initial discriminability, *a*, but did not have a systematic impact on the rate of decay, *b*, as found in the present experiment. Thus, the results of the present study provide support for the findings of Adamson (1995). This finding is also similar to that of White (1985) who found that increasing the discriminability between two sample stimuli results in a change in *a* but no systematic change in the rate of decay *b*.

The previous experiment suggested that there was little or no difference in how well the simple negative exponential (Equation 1) and the negative exponential (sqrt(*t*)) (Equation 2) fitted the data. The present result also suggests that the fits provided by these two functions do not differ markedly.

Given that neither the present data or the data from Experiment 1 support the use of one function over the other, in terms of goodness of fit, the decision to use Equation 2 (negative exponential sqrt(*t*)) was made on the basis of the theoretical justifications given by White (2001), White et al. (1996) and Wixted and Ebbesen (1991). Given this, the negative exponential (sqrt(*t*)) (Equation 2) will be used to describe the data in the remaining experiments here, and this function will simply be referred to as the negative exponential.

Overall, the present result suggests that increasing the number of sample stimuli used in a DMTS task results in an increase in the physical disparity

reached at the end of the titration procedure. The present results also suggest that increasing the number of sample stimuli used in a DMTS gives a decrease in accuracy. Thus, it was argued that when the intention is to investigate the impact of increasing the number of sample stimuli on matching accuracy, the titration should use the smallest number of sample stimuli to be used in the DMTS. This would then allow high accuracy for the smallest sample set and accuracy could also decrease as the number of sample stimuli was increased.

Experiment 3

The previous two experiments suggested that increasing the number of sample stimuli used in a titration resulted in an increase in the physical disparity required to reach a pre-set level of accuracy. However, given that not all of the participants reached the criterion by the end of all of the titrations these results were not conclusive. As none of the participants had reached the criterion for the 16-sample titration, in Experiment 2, and few had reached the criterion for the 2-sample titration, in Experiment 1, it was not clear if it was possible for a participant to reach the pre-set criterion at all for larger sample sets.

The present experiment aimed to investigate whether the physical disparity required to meet a pre-set criterion level of accuracy (100%) increased systematically as the number of sample stimuli used during the titration was increased, across a range of sample-set sizes. In the present experiment there were four titrations with 1, 4, 8 and 16 sample stimuli. Given the problems in reaching the pre-set criterion, encountered in Experiments 1 and 2, the maximum allowable physical disparity was increased to 50%. The maximum number of trials was also increased from 200 to 500 for each titration to allow more trials to reach the larger physical disparities.

Method

Participants

Five first-year psychology students participated in this experiment. Each participant received course credit towards one of two first year psychology courses, irrespective of their performance on the experimental task. The data of one of the participants was not used due to an equipment malfunction.

Apparatus

The apparatus used was the same as that used in Experiment 1.

Stimuli

Filled white disks of varying size, presented on a blue background, were used as stimuli. The sample stimuli were drawn from a pool of 16 disks ranging in size from 46 pixels to 151 pixels in diameter, in steps of 6 pixels. There were 4 different sized sample sets with 1, 4, 8, and 16 stimuli. When there was one sample stimulus it was 95 pixels in diameter. The 4, 8 and 16 sample-stimulus sets contained, respectively, all the stimuli between and inclusive of 88 to 109, 74 to 123 and 46 to 151 pixels in diameter. The non-matching comparison stimuli differed in size from the sample stimuli by a proportion of the diameter of the sample stimulus. For each sample stimulus there were two possible non-matching sample stimuli, one larger and one smaller than the sample stimulus.

Procedure

The general procedure for the present experiment was the same as for the one and the two-sample titrations used in Experiment 1 except for the details given below. In the first titration there were 16 sample stimuli, in the second titration there were 8, in the third titration there were 4 and in the fourth titration there was 1 sample stimulus. There was a total of 500 trials allowed for each of the titrations and the maximum physical disparity that could be reached was 50% of the diameter of the sample stimulus. For each condition, the physical disparity between the sample stimuli and the non-matching comparison stimuli was initially set at 4%. Each condition ended when a participant was 100% correct across 30 consecutive trials at one physical disparity, or when a participant had completed the maximum of 500 trials, or when the maximum physical disparity of 50% was reached. The experiment ended at the completion of Condition 4. Once the experimental session had ended the participants were given information about the purpose of the experiment.

Results

Figure 3.1 shows correct and incorrect responses, as a function of trial number and physical disparity for the 1, 4, 8 and 16 sample titrations for the each participant as previously plotted in Figure 1.2. All of the participants reached the criterion of 100% correct across 30 trials, for all of the titration conditions. The figure shows that at the start of each titration when the physical disparity was small the slopes of the graphs were generally steep, for all of the titrations and for all of the participants. This indicates that all of the participants were making a large number of errors. As the physical disparity increased the slope of the graphs become shallower, indicating that the participants were making fewer errors. With the exception of the 16-sample titration for S1 and the 1-sample titration for S3 and S4 the physical disparity at which the participants first completed 30 trials at one physical disparity, was not the one at which they finally met the pre-set criterion. This is shown by the flat step-like sections of the graphs where only a few errors were made.

Table 3.1 shows the total number of trials that each of the participants took to complete each of the four titrations. None took the maximum number of 500 trials to reach criterion for any of the titration conditions. For S2 and S3 all of the titration conditions were completed within 200 trials. All of the titration conditions for S1 and S4 took more than 200 trials to complete, with the exception of the 16-sample titration for S1 and the 1-sample titration for S4. The smallest number of trials taken to complete a titration condition was 90 trials for S4 for the 1-sample titration. The largest number of trials taken to complete a titration condition was 350 trials for S1 for the 4-sample titration condition.

Figure 3.2 and Table 3.1 show the physical disparity reached at the end of each of the titration conditions for each individual and the average of these. None of the participants reached the maximum physical disparity of 50% for any of the titration conditions. For S1 and S4 the physical disparities reached for all of the titrations were generally larger than those reached for S2 and S3.

There was no systematic increase in the physical disparity at which the criterion was reached as the number of sample stimuli used for the titration was

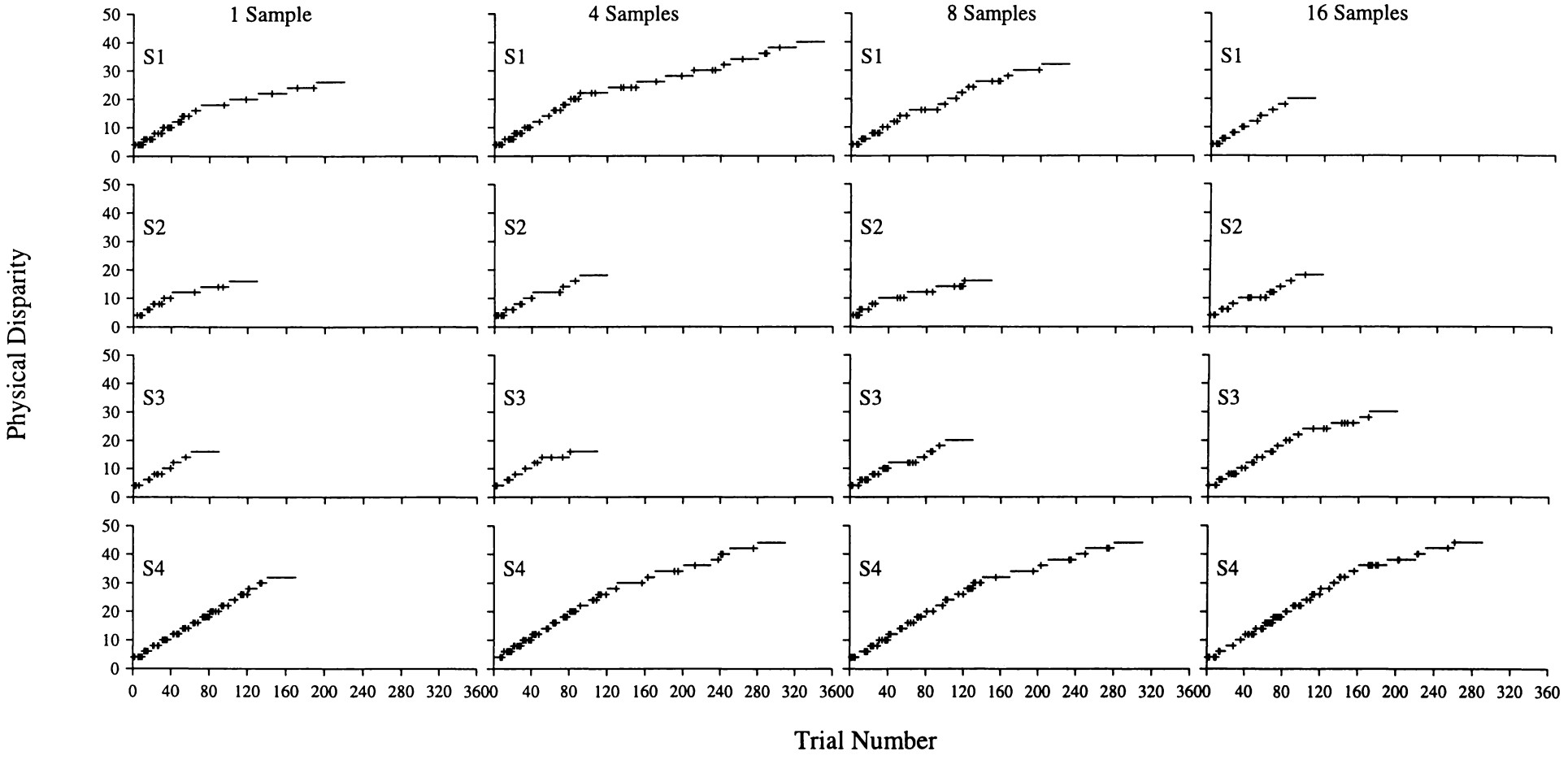

Figure 3.1. Correct (dot) and incorrect (plus) responses, for each titration, as a function of experimental trial and physical disparity for each of the individual participants.

Table 3.1 *The physical disparities at which the individual participants first obtained 100% correct across 10 consecutive trials and the physical disparity at which they reached the criterion, for each of the conditions. Also shown is the total number of trials each of the participants took to complete each titration. The average across all participants is also shown for all these measures.*

		Physical disparity			
		1 sample	4 samples	8 samples	16 samples
S1	10 trials	18	22	16	20
	30 trials	26	40	24	20
	Total Trials	220	350	230	110
S2	10 trials	12	12	10	10
	30 trials	16	18	16	18
	Total Trials	130	120	150	120
S3	10 trials	16	14	12	24
	30 trials	16	16	20	30
	Total Trials	90	110	130	200
S4	10 trials	32	30	32	36
	30 trials	32	44	44	44
	Total Trials	170	310	310	290
Average	10 trials	19.5	19.5	17.5	22.5
	30 trials	22.5	29.5	26	28
	Total Trials	153	223	205	180

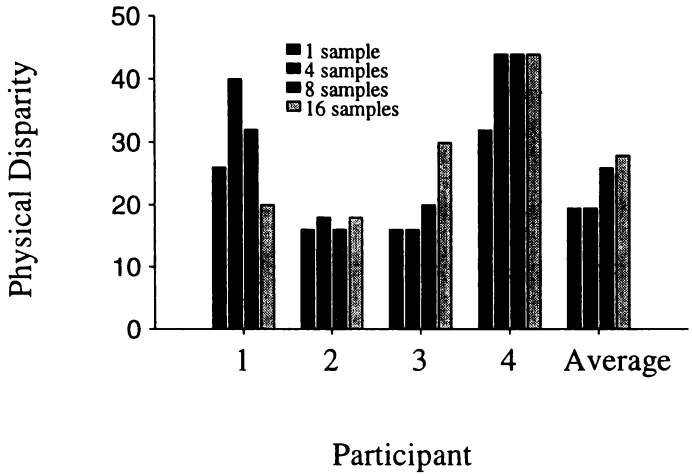

Figure 3.2. Physical disparities for each of the four Participants for each of the four titration procedures in which the number of sample stimuli was varied.

increased for any of the participants. For S3 the physical disparity increased systematically as the number of sample stimuli was increased from 4 to 8 and from 8 to 16. For S1 there was a systematic decrease in the physical disparity as the number of sample used was increased from 4 to 8 to 16. The physical disparity reached for the 4-sample titration and for the 16-sample titration was larger than the physical disparity reached for the 1-sample titration for three of the four participants. For the eight-sample titration the physical disparity reached was larger than that reached for the one-sample titration for two out of the four participants. There was no difference in the physical disparity reached at the end of the 1- and the 4-sample titration for the averaged data. But the physical disparity reached for the averaged data increased systematically as the number of sample stimuli used for the titration was increased from 4 to 8 and to 16.

The data from individuals was also analysed to see at what physical disparity each of the participants first obtained 100% correct across 10 trials and so continued for a further 20 trials at the same physical disparity. These data are shown in Table 3.1. There was no systematic change in the physical disparity as the sample-set size increased for the first set of ten trials at which a participant was first 100% correct.

Discussion

The present experiment showed that increasing the maximum number of trials and the maximum physical disparity allowed all of the participants to reach the criterion for all of the sample-set sizes. However, there was no systematic increase in the physical disparity reached at the end of the titration as the number of sample stimuli was increased.

All participants completed the titration for all of the sample-set sizes within 500 trials, but the number of trials required varied across the participants. Given that 350 was the largest number of trials required the present result suggests that it was not necessary to increase the maximum number of trials to 500. The maximum physical disparity, between the sample and the non-matching comparison stimuli, was increased to 50% of the diameter of the sample stimulus.

However, the largest physical disparity reached for any of the participants was 44%. Although this result suggested that increasing the maximum physical disparity allowed participants to complete the titration procedures, it also suggested that the maximum physical disparity could have been less than 50%.

Although sample-set size did not relate systematically to the final physical disparity, the physical disparity reached at the completion of the 16-sample titration was greater than that reached at the completion of the 1-sample titration for three of the four participants. This result is consistent with the finding of the previous experiment, where the physical disparity reached for the 16-sample titration was generally larger than that reached for the 1-sample titration. In the present experiment all participants successfully completed both the 1 and the 16-sample titrations and, therefore, the physical disparities obtained were for the same level of accuracy. These data support the previous suggestion that when investigating the impact that increasing the number of sample stimuli has on matching accuracy for a DMTS task, the smallest sample-set should be used for the titration. This would allow initial discriminability to be high for the smallest sample set size and to decrease as the number of sample stimuli was increased which would be less likely lead to a ceiling effect.

Experiment 4

One feature of the procedure used in the previous three experiments is that participants were not given feedback about whether their responses were correct or incorrect. Typically when animal memory is studied correct responses result in some sort of consequence, such as food or water (termed reinforcers), to maintain responding and or to give feedback on the correctness of the response. However, when human memory is studied such consequences are less common.

When using humans participants some researchers have used consequences similar to those typically used with animals. Forzano and Logue (1995), for example, used food and drink with human participants. Typically when such consequences are used with animals a state of deprivation is induced prior to the experiment. However, as Galizio and Buskist (1988) point out, it is hard to bring about or control deprivation with human participants and, because of this they suggested that food and drink may not always function as effective consequences when used with adult humans (Galizio & Buskist, 1988; Weardon, 1988). For example, Buskist (cited in Galizio and Buskist, 1988) and Miller (cited in Weardon 1988) found that when food was used as an experimental consequence that participants frequently failed to consume and often discarded the food. Given these problems the use of these types of consequences with humans has not been frequent.

A variety of other consequences have also been used with human participants. Sometimes these consequences are accumulated during the session. For example, participants have accumulated money (Buskit & Miller, 1986; Stromer, McIlvane, Dube & MacKay, 1993; Bowden, Benedikt & Ritter, 1992), points (Spiga, Cherek, Grabowski & Bennet, 1992) or tokens (Newman, Buffington & Hemmes, 1995) during the session. Sometimes these points and tokens are exchanged for commodities such as lottery tickets, food or money after the end of the experimental session (Wurster & Griffiths, 1979; Bardenhagen & Bowden, 1995; Newman, Buffington & Hemmes, 1995). Galizio and Buskist (1988) and Weardon (1988) report that points or points exchangeable for money have been most commonly used with human participants. They report that points

do not functions as an effective consequence in all cases. However, Ellis (1978) and Weardon (1988) suggested that while reinforcers may not always function to strengthen responding in human participants experimental consequences, such as points, may function as a source of information. That is, consequences may provide human participants with information about whether their responses were correct or not and thus, be used to direct performance.

Feedback other than points has also been used with humans. Wilson and Hayes (1996), provided human participants with visual feedback on the accuracy of their responding while Trigo, Martinez and Moreno (1995) provided their participants with verbal feedback. Stromer, McIlvane, Dube and MacKay (1993) and Holdstock, Shaw and Aggleton (1995) used both visual and auditory feedback. There are a few studies that have investigated whether such feedback has an effect on human performance. For example, Trigo, Martinez and Moreno (1995) and Wilson and Hayes (1996) both investigated whether participant's responding was effected by the removal of feedback. Both studies found that their participants continued to respond when feedback was withdrawn. Additionally, Tarr, Kersten and Bülthoff, (1998) found that feedback had no effect on the performance of their human participants across the experimental session. Thus, it is not clear whether either direct reinforcement within the session or feedback are needed, or effective, when human participants are to be used for experimental studies.

If feedback provides participants with information about their accuracy on a task and allows improved performance then it is possible that with feedback performance may change across the experimental session. None of the previous experiments, here, gave a direct consequence of any kind following a response and there was no evidence of any change in accuracy across the experimental session. However, given that consequences can result in improved accuracy with animals it is possible that accuracy might change over the session for the present task if consequences were given.

The aim of the present experiment was to investigate the effect of feedback on the matching accuracy of human participants. The procedure used for the present study was similar to that used for the previous studies, here. The

initial part of the experimental session was a titration in which participants were not given feedback. This was followed by a DMTS task using a repeated-condition within-subject design. As it was not clear what effect feedback would have on accuracy so the pre-set criterion for the titration was set to 70% correct to allow room for accuracy to increase with feedback. A beeping sound produced by the computer was given following a correct response, as feedback, in the Feedback Condition.

Method

Participants

Five first year psychology students participated in this experiment. Each participant received 1% course credit towards a first year psychology course. Course credit was awarded irrespective of the participant's performance and irrespective of whether they completed the experiment.

Apparatus

The apparatus was the same as that used in Experiment 1.

Stimuli

Two white disks, 80 and 83 pixels in size were used as sample stimuli. They were presented on a blue background. For each sample stimulus there were two non-matching comparison stimuli, one larger and one smaller than the sample stimulus. The non-matching comparison stimuli differed in size from the sample stimulus by a proportion of the diameter of the sample stimulus. The exact proportional difference was selected during the initial part of the experiment, for each participant.

Procedure

The general procedure used was the same as that used for Experiment 1. The instructions for the present experiment differed from those given to the participants for Experiment 1 in that the participants were also informed that sometimes during the experimental procedure they would hear a beep when they

had made a correct response. The general procedure also differed from that of Experiment 1 in that after the participants had read the instructions the experimenter caused the computer to produce the beep that was to be used as feedback to each of the participants.

The initial part of the session involved a two-sample titration procedure similar to that used for Experiment 2. The present experiment differed in that the pre-set criterion level of accuracy for the present experiment was set to 70% correct across 30 trials. As for Experiment 2 the total number of trials for the titration was 200 and the maximum physical disparity was 28%. The main part of the session involved a DMTS task, with the same delays as used in Experiment 2. There were three experimental conditions and Conditions 1 and 3 were the same as in previous experiments in that participants were not informed whether their responses were correct or incorrect. During Condition 2 each correct response was followed by a beep, which was produced by the computer.

For each participant the stimulus disparity was the same across all three conditions and was set at the physical disparity reached at the end of the titration, for that participant. There were a total of 96 trials in each of the conditions, with 32 trials at each of the delay intervals. The experimental trials were pre-programmed and so the order in which the experimental events occurred was the same for all of the participants. The delay intervals were presented in the following order, 4, 16, 8 and 0.05 s in blocks of 32 trials, in all three conditions. The sample stimulus for each trial was pre-selected using a random number generator with the following constraints, the number of times each of the two sample stimuli was presented must be equal and each of the sample stimuli must not be presented more than three times on consecutive trials. Each sample stimulus was presented an equal number of times with its larger and its smaller non-matching comparison stimulus. Each condition ended after the completion of 96 trials and the next condition began immediately. The experimental session ended when the participant had completed 288 trials at the conclusion of Condition 3. Once the experimental session had ended the participants were given information about the purpose of the experiment.

Results

Unfortunately it was discovered on detailed analysis of the results that the titration task had not functioned correctly. The titration ended before the maximum number of trials had been completed but none of the participants had completed 30 trials with the physical disparity which they had reached when the titration ended. In addition for all but S3 the program increased the physical disparity by one step at the start of the DMTS phase of the session. The physical disparities for S1 to S5 used in the DMTS task were 11, 11, 9, 11 and 13% correct, respectively. It was decided to use these data as, although the titration did not function as planned, the participants were generally performing at or above 70% at the end of the titration and, with the exception of S2, all performed at 70% or above with the zero delay in the DMTS task.

Figure 4.1 shows percent correct for each of the five participants and the averaged data plotted for each delay and each experimental condition. There was no systematic change in accuracy as a function of whether feedback was given for correct matching responses, for any of the participants at any delay interval. The average data are consistent with the data of the individual participants in that there was no systematic change in accuracy at any delay as a function of whether or not feedback was given.

Figure 4.2 shows logit *p*, calculated using Equation 3, for each of the five participants and the averaged data for each delay interval and experimental condition. There was no systematic change in matching accuracy, for any of the participants or for the averaged data, as a function of whether feedback was given for any delay. A repeated measures ANOVA was carried out on these data. Prior to the ANOVA a Mauchly's test of sphericity was carried out to test for homogeneity of variance. The main effects of feedback ($W = 0.15$) and delay ($W = 0.30$) for this test were not significant but the interaction ($W = 0.00$) was significant ($\alpha > 0.05$). As a result the Huynh-Feldt correction to the degrees of freedom was used in the main ANOVA. For the ANOVA the main effects of feedback ($F = 0.11$, $df = 1.17, 4.68$), delay ($F = 2.24$, $df = 3.00, 12.00$) and the interaction ($F = 0.81$, $df = 6.00, 24.00$) were not significant ($\alpha < 0.05$).

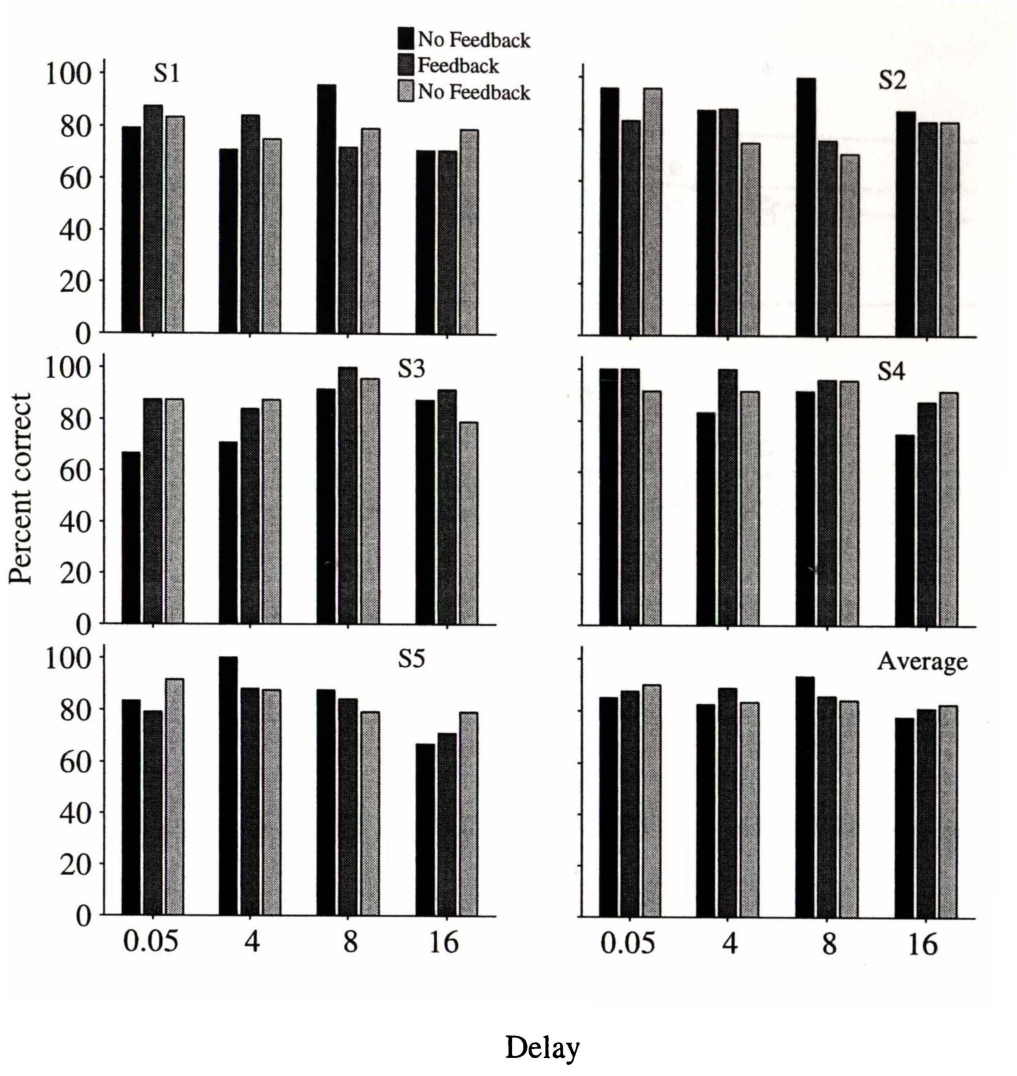

Figure 4.1. Percent correct as a function of whether feedback was provided for each delay interval for each participant and the average of these.

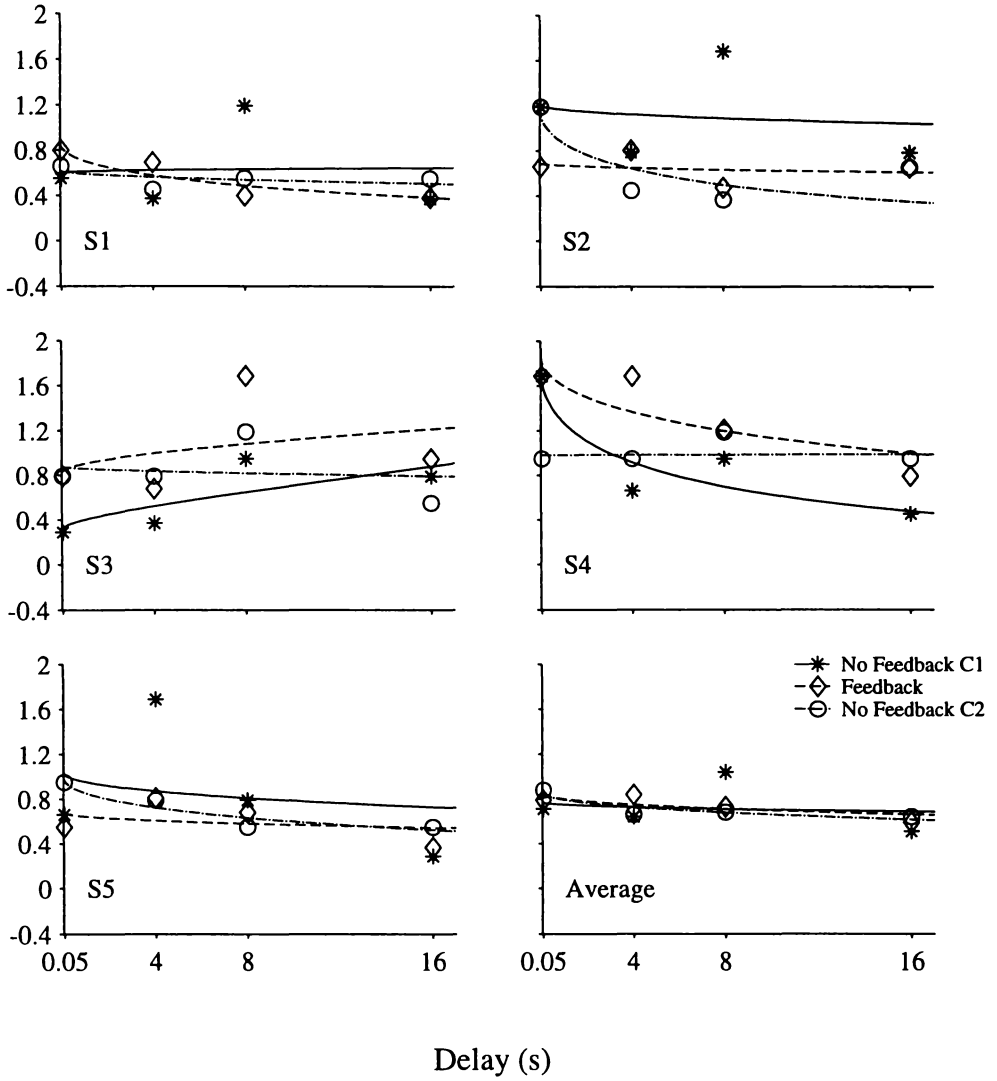

Figure 4.2. Logit *p* plotted as a function of delay and feedback for each participant and the average of these. The fitted functions are negative exponentials (Equation 2).

Negative exponential functions (Equation 2) were fitted to the logit *p* data using non-linear regression and are shown on Figure 4.2. Table 4.1 shows the values obtained for the parameters *a* and *b* for both the individual and the averaged data. Table 4.1 also shows the standard errors of estimate and the percentages of variance accounted for (VAC) by the fitted functions. For the individual participants *a* changed across the three conditions but these changes were not systematic across the participants. For S1 and S4, *a* was higher when there was feedback (Condition 2), than when there was not (Condition 1 and Condition 3). For S2 and S5, *a* was lower when there was feedback (Condition 2) than when there was not (Condition 1 and Condition 2). For S3 and for the averaged data *a* increased systematically across Conditions 1 to 3. Accuracy tended to decrease slightly as a function of delay, for both the individual and the averaged data, as shown by the *b* values in Table 4.1. Although for some participants there were negative *b* values, which indicated that matching accuracy increased as a function of delay. There was no systematic change in *b* as a function of experimental condition for either the individual or the averaged data.

An analysis of response bias based on the size of the comparison stimuli was carried out, for each of the participants as a function of delay and whether or not feedback was given. This bias was calculated as the logarithm of the ratio of the total number of times that the small comparison stimulus was selected to the total number of times that the large comparison stimulus was selected. These data are shown in Figure 4.3. There was little or no bias for any participant or for the average as a function of the size of the comparison stimulus and this bias did not change as a function of whether or not feedback was given.

The individual data were also analysed for any bias, as a function of whether feedback was given and delay, based on the side of the screen on which the comparison stimulus appeared on. These data are shown in Figure 4.4. There was no systematic bias towards selecting either the left or the right comparison stimulus for any of the participants or for the average.

The data for each of the participants was analysed for any changes in performance across the experimental session and are shown in Figure 4.5. Data from each of the three conditions was divided into two blocks of 16 trials for each

Table 4.1 *Estimates of a, b, (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions for each of the experimental conditions for each individual and the average of these.*

	Condition	*a*	*b*	VAC%	Std Err
S1	No feedback	0.60	-0.02	0	0.34
	Feedback	0.87	0.21	84	0.07
	No feedback	0.62	0.05	27	0.06
S2	No feedback	1.21	0.04	2	0.37
	Feedback	0.69	0.03	5	0.11
	No feedback	1.19	0.30	62	0.20
S3	No feedback	0.31	-0.26	59	0.18
	Feedback	0.82	-0.10	13	0.37
	No feedback	0.88	0.03	2	0.23
S4	No feedback	1.77	0.33	85	0.18
	Feedback	1.88	0.16	71	0.20
	No feedback	0.98	-0.02	5	0.10
S5	No feedback	1.04	0.09	6	0.50
	Feedback	0.68	0.56	10	0.15
	No feedback	1.00	0.16	90	0.05
Average	No Feedback	0.77	0.03	3	0.19
	Feedback	0.85	0.06	61	0.06
	No feedback	0.86	0.08	80	0.04

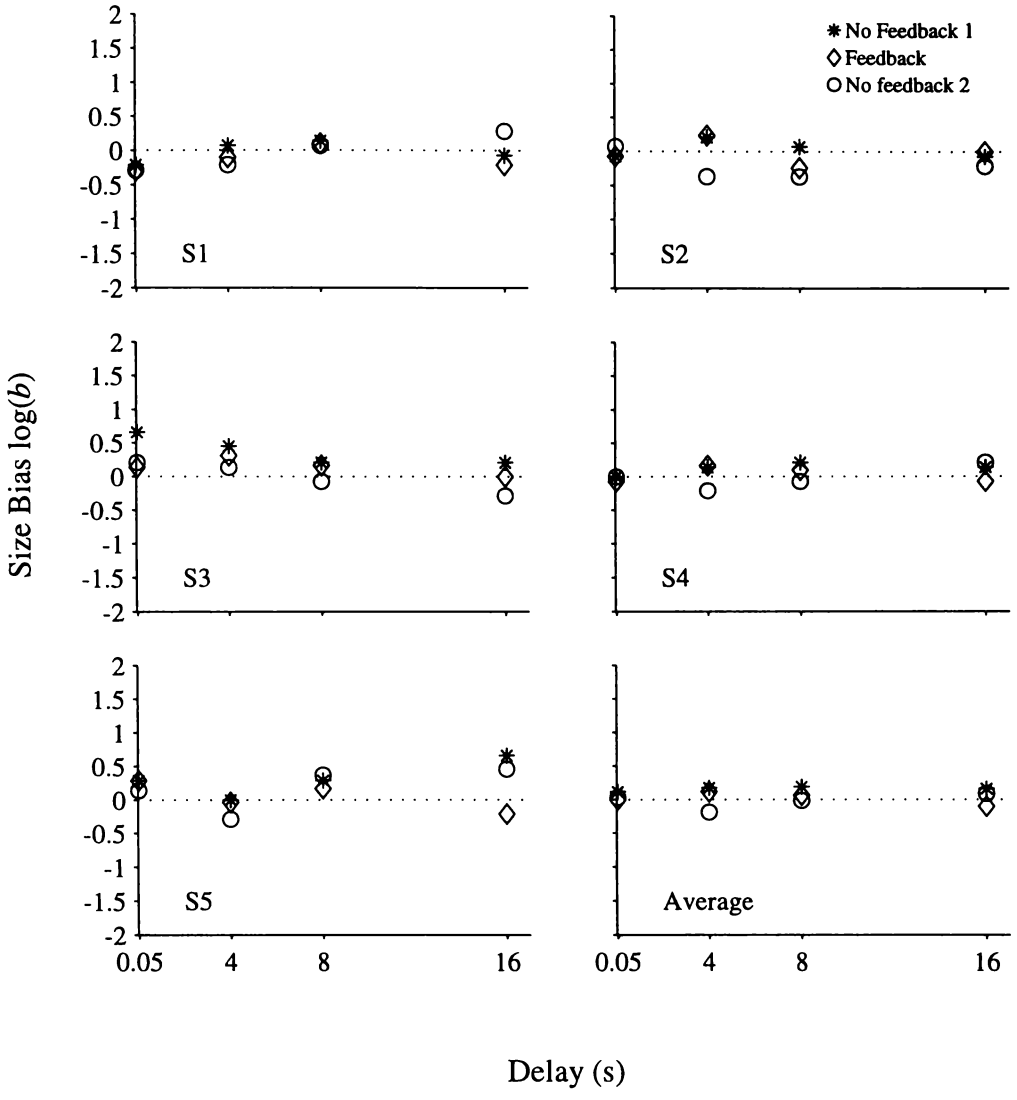

Figure 4.3. Log response bias plotted as a function of the size of the comparison stimuli for each participant and the average of these.

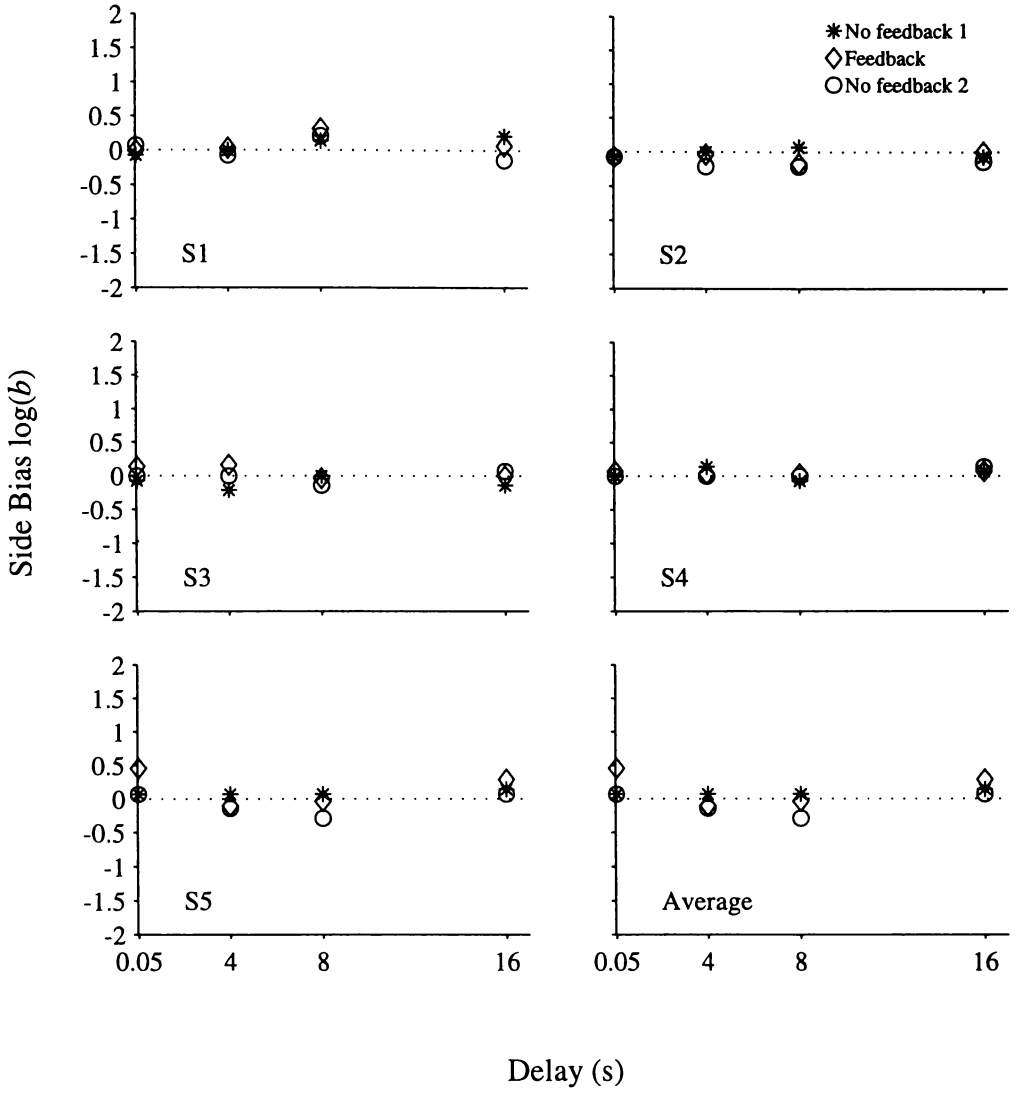

Figure 4.4. Log response bias plotted as a function of the side of the screen on which the comparison stimuli appeared, for each participant and and the average of these.

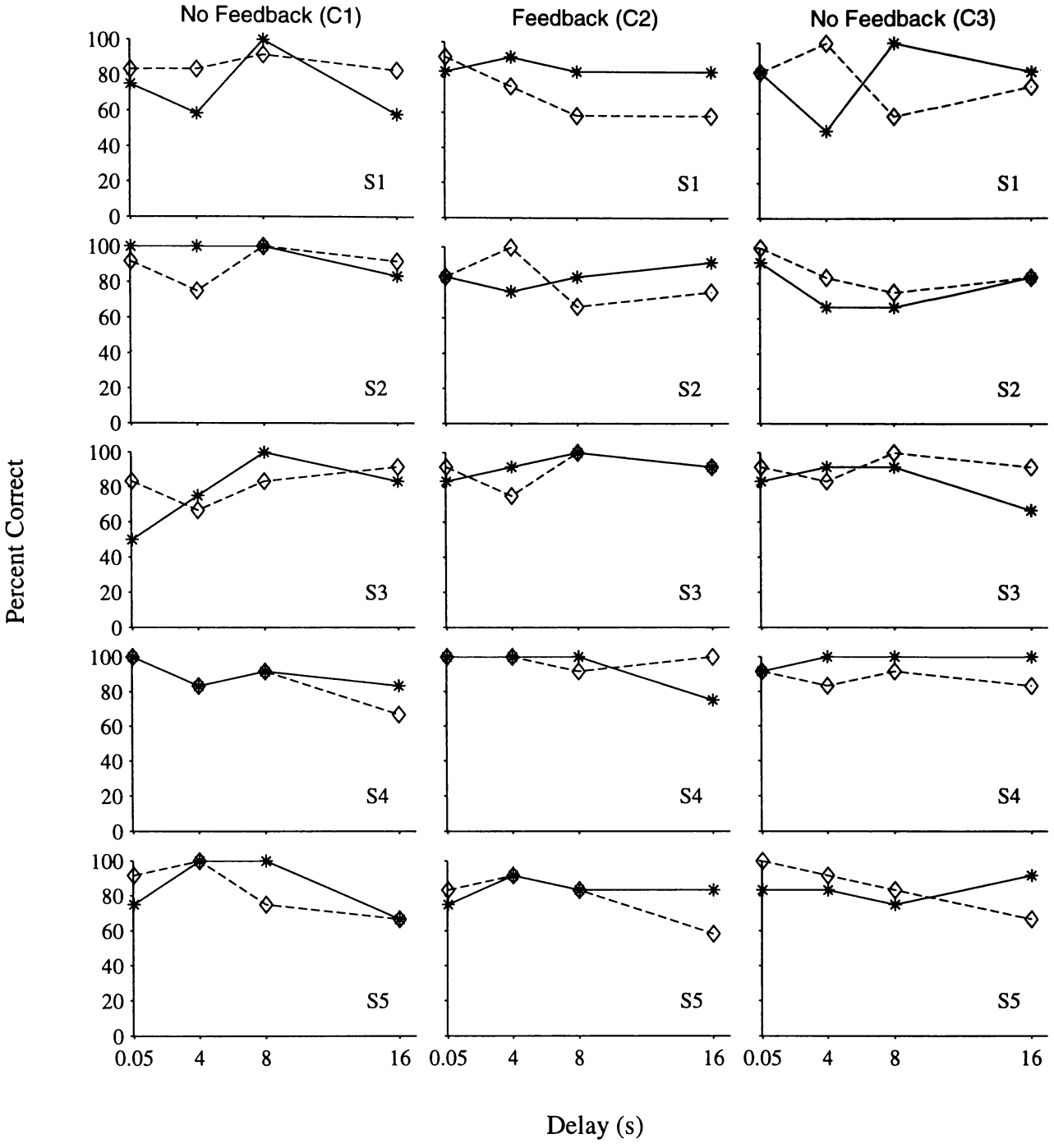

Figure 4.5. Percent correct on the first and second block of 16 trials for each experimental condition for each participant plotted as a function of delay (see text for details).

of the delay intervals. The data were then examined for any systematic change in accuracy, across the two blocks of 16 trials, as a function of delay interval and whether or not feedback was given. There were no systematic changes in accuracy across the two blocks for any of the participants.

Discussion

The addition of feedback in Condition 2 had no systematic effect on accuracy across the participants. The present result also showed that *a* and *b* did not change systematically as a function of whether feedback was given for correct responses across participants.

The finding that feedback had no effect when withdrawn here is consistent with the results of Trigo et al. (1995) and Wilson and Hayes (1996). Thus, the present study suggests that feedback does not change the performance of human participants in the present task. The present experiment is also consistent with the finding of Tarr et al. (1998) who found the feedback did not result in any change in performance across the experimental session.

In the present experiment the pre-set criterion for the titration was set to 70% to allow room for accuracy to increase with feedback. However, there was generally no change in accuracy, here, when feedback was given. Thus, given that feedback had no systematic effect on matching accuracy for the task used here, feedback was not be used in the following experiments

Experiment 5[1]

The previous experiments established that a titration could be used, in a DMTS task, to produce similar levels of performance across participants and that the addition of feedback did not affect accuracy. With this established, it was possible to start investigating the original experimental question which was to examine the way in which the number of sample stimuli used in a DMTS task effects performance. As mentioned earlier, a number of previous studies have examined the impact of sample number on matching accuracy. These studies have lead to three different findings. These were that as the sample-set size was increased; first, that matching accuracy did not change (Etkin and D'Amato, 1969; Mason and Wilson, 1974), second, that matching accuracy increased (Worsham, 1975; Mishkin and Delacour, 1975) and, third, that matching accuracy decreased (Roberts, 1980; and Adamson, 1995). The aim of the present experiment was to establish the effect of increasing the number of sample stimuli used in the present DMTS task on matching accuracy.

Although the number of sample stimuli was varied across the DMTS task, used here, only a single titration was used and only one sample was used in the titration. This decision was made because the number of sample stimuli was to be varied in the DMTS task, and the intention was to compare performance across these different sample-set sizes. Thus, the titration procedure used the smallest sample-set size, which was, here, one sample stimulus.

There are a number of advantages to using a one-sample titration. First, it generally takes fewer trials with one-sample than with more that one sample stimulus for participants to reach the pre-set criterion level of accuracy. Second, the results of Experiment 2 suggested that when the number of samples was varied for the DMTS task used here accuracy decreased as the number of sample stimuli was increased. Thus, using a one-sample titration would mean that accuracy is high for the smallest sample set-size and that any decreases which may occur as

---

[1] The data from this experiment have been published.
Adamson, C., Foster, T.M., & McEwan, J. S.A. (2000). Delayed matching to sample: the effects of sample-set size on human performance. *Behavioural processes, 49,*149-161.

the number of sample stimuli was increased, or as the delay interval was increased, will be observable.

In the present study participants were exposed to a series of conditions in which the number of sample stimuli used was increased across a range of delays. The sample-set sizes used were 1, 4, 8, 16 and 32. This range of sample-set sizes was chosen to encompass the number of samples used in studies mentioned earlier.

Method

Subjects

Six first year psychology students participated in this experiment. Each participant received course credit towards one of two first year psychology courses, irrespective of their performance on the experimental task.

Apparatus

The apparatus used was the same as that used in Experiment 1.

Stimuli

Filled white disks of varying sizes, presented on a blue background, were used as stimuli. The sample stimuli were drawn from a pool of 32 disks ranging in size from 42 pixels to 135 pixels in diameter, in steps of 3 pixels. There were 5 different sized sample sets with 1, 4, 8, 16 or 32 stimuli. When there was one sample stimulus it was 87 pixels in diameter. The 4, 8 16 and 32 sample-stimulus sets contained, respectively, all the stimuli between and inclusive of 84 to 93, 78 to 99, 66 to 111 and 42 to 135 pixels in diameter as shown in Table 5.1. The non-matching comparison stimuli differed in size from the sample stimuli by a proportion of the diameter of the sample stimulus, as derived by the initial titration. For each sample stimulus there were two possible non-matching sample stimuli, one larger and one smaller than the sample stimulus.

Table 5.1. *The size of the stimuli pixels in diameter, in the sample pool and the sample-set size where the sample appeared as a stimulus.*

Sample size	Sample-set size where the sample appeared as a stimulus
42	32
45	32
48	32
51	32
54	32
57	32
60	32
63	31
66	16, 32
69	16, 32
72	16, 32
75	16, 32
78	8, 16, 32
81	8, 16, 32
84	4, 8, 16, 32
87	1, 4, 8, 16, 32
90	4, 8, 16, 32
93	4, 8, 16, 32
96	8, 16, 32
99	8, 16, 32
102	16, 32
105	16, 32
108	16, 32
111	16, 32
114	32
117	32
120	32
123	32
126	32
129	32
132	32
135	32

Procedure

The general procedure was the same as that used for Experiment 2. The initial part of the session was a one sample titration. A single sample stimulus, 87 pixels in diameter, was used during this phase of the experiment. The titration differed from that used in Experiment 2 in that it ended if the participant reached a proportional difference of 30% and had not achieved the 100% criterion. If this occurred, the main part of the session began using the maximum proportional difference of 30%.

The main part of the session involved a DMTS task. The general procedure was the same as that used for Experiment 2 except that there were four delays (0.5, 4, 8 or 16 s) between the presentation of the sample stimulus and the choice phase. The DMTS procedure also differed in that there were five conditions each with a different sample set size. There were 32 sample stimuli in Condition 1, 16 sample stimuli in Condition 2, 8 sample stimuli in Condition 3, 4 sample stimuli in Condition 4, and 1 sample stimulus in Condition 5. There was a total of 160 trials in each of the conditions, with 32 trials at each of the delay intervals and the next condition began immediately. The experimental session ended when the participant had completed 800 trials at the conclusion of Condition 5. Once the experimental session had ended the participants were given details about the purpose of the experiment.

Results

S1 to S6 completed the titration in 120, 130, 70, 150, 120 and 110 trials respectively and all of the participants were performing at the pre-set level of accuracy by the end of the titration. The proportional disparities reached were 14, 30, 12, 16, 18 and 16% for S1 to S6 respectively.

Figure 5.1 shows percent correct for each of the six individual participants and the average of these as a function of delay for each of the different sample set sizes. For the individual data accuracy decreased slightly as the delay was increased and there was a tendency for matching accuracy to change as the number of sample stimuli was increased. With the exception of S3, accuracy at

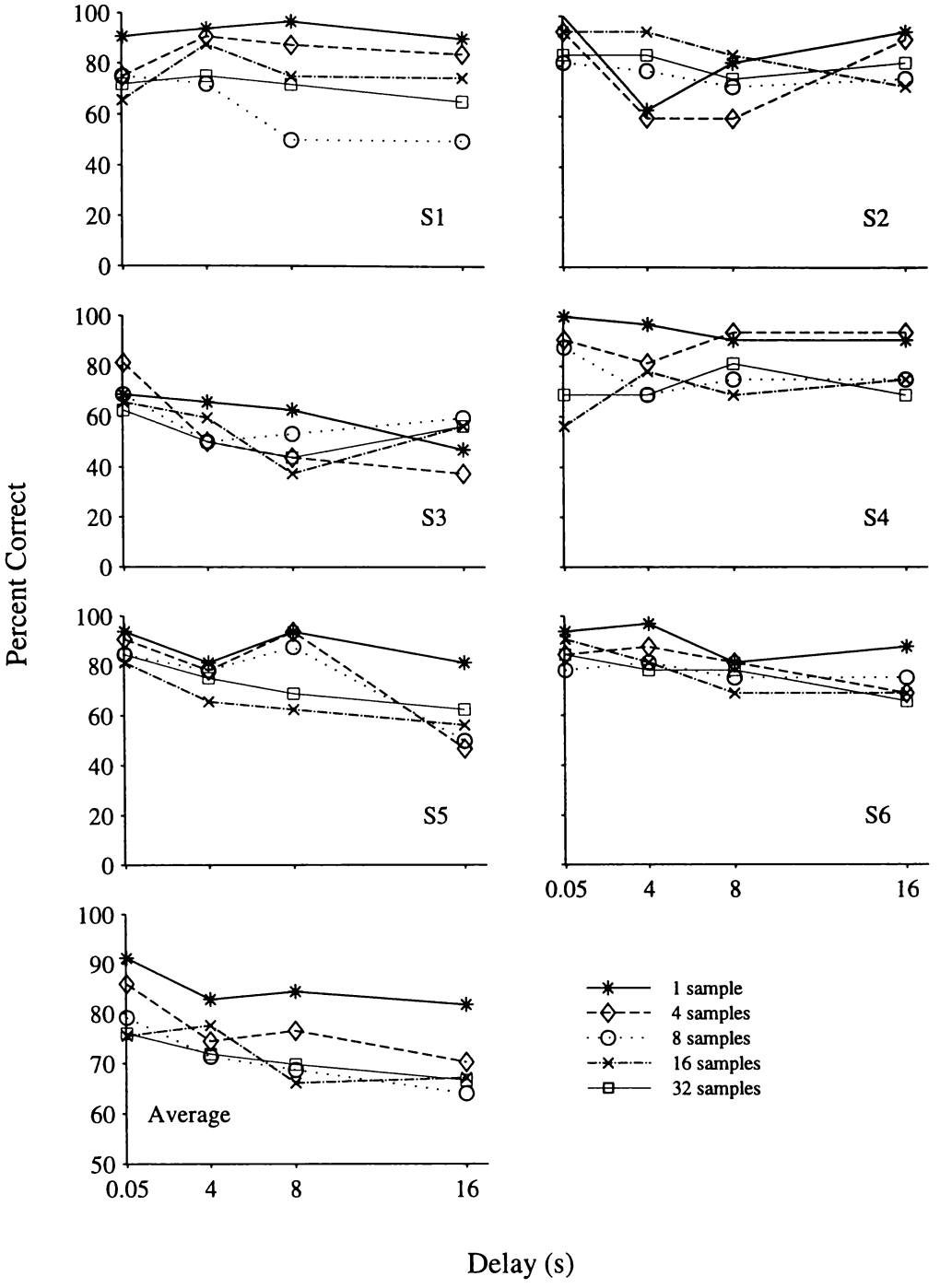

Figure 5.1. Percent correct plotted as a function of delay and sample-set size for each participant and the average of these.

0.05 s delay was highest when there was one sample stimulus and tended to decrease as the number of sample stimuli was increased. However, it is difficult to see in the individual percent correct graphs these changes in matching accuracy. For the average data accuracy decreased slightly as the delay interval was increased. Matching accuracy was highest, across all delays, when there was one sample stimulus. When there were 4 sample stimuli matching accuracy was higher for all delays than when there were 8, 16 and 32 sample stimuli, with the exception of the 4 s delay. There was no notable change in matching accuracy when the number of sample stimuli was increased from 8 to 16 and from 16 to 32 sample stimuli.

In order to examine individual changes in matching accuracy with sample number, percent correct, at each delay, was ranked from highest to lowest across sample-set size. Table 5.2 shows these rank orders of the percentages correct for each of the sample-set sizes, averaged across delay interval for each of the individual participants. With the exception of one condition, rank order systematically decreased as the sample-set size was increased, for four out of the six participants. With the exception of two conditions, rank order decreased systematically as the sample set size was increased for the remaining two participants.

Figure 5.2 shows logit p, calculated using Equation 3, for each of the six individual participants and the average of these as a function of delay for each of the sample set sizes. In two cases there were zero errors for individual data. As for the percent correct data the individual data logit p tended to decrease as the number of sample stimuli increased. For the average data logit p was clearly higher for all delays when there was only one sample stimulus (Condition 5), than for all other conditions. Logit p when there were four sample stimuli (Condition 4), was higher for all delays than for Conditions 3, 2 and 1 with 8, 16 and 32 sample stimuli respectively. There was no consistent change in logit p at zero delay, with 8, 16 and 32 sample stimuli respectively (Conditions 3, 2 and 1 respectively).

A repeated measures ANOVA was carried out on these data. Prior to the ANOVA Mauchly's test of sphericity was used to test for homogeneity of

Table 5.2 *Rank order of percent correct performance for each of the sample set sizes, averaged across delay interval.*

	Average rank across delays				
	Sample number				
	1	4	8	16	32
S1	1.00	2.13	4.38	3.50	4.00
S2	2.00	3.63	4.00	2.63	3.00
S3	1.88	3.13	2.38	3.38	3.75
S4	1.50	1.50	3.00	4.13	4.13
S5	1.13	2.75	3.13	4.25	3.25
S6	1.13	2.75	3.63	3.50	4.00
Average	1.44	2.63	3.42	3.54	3.69

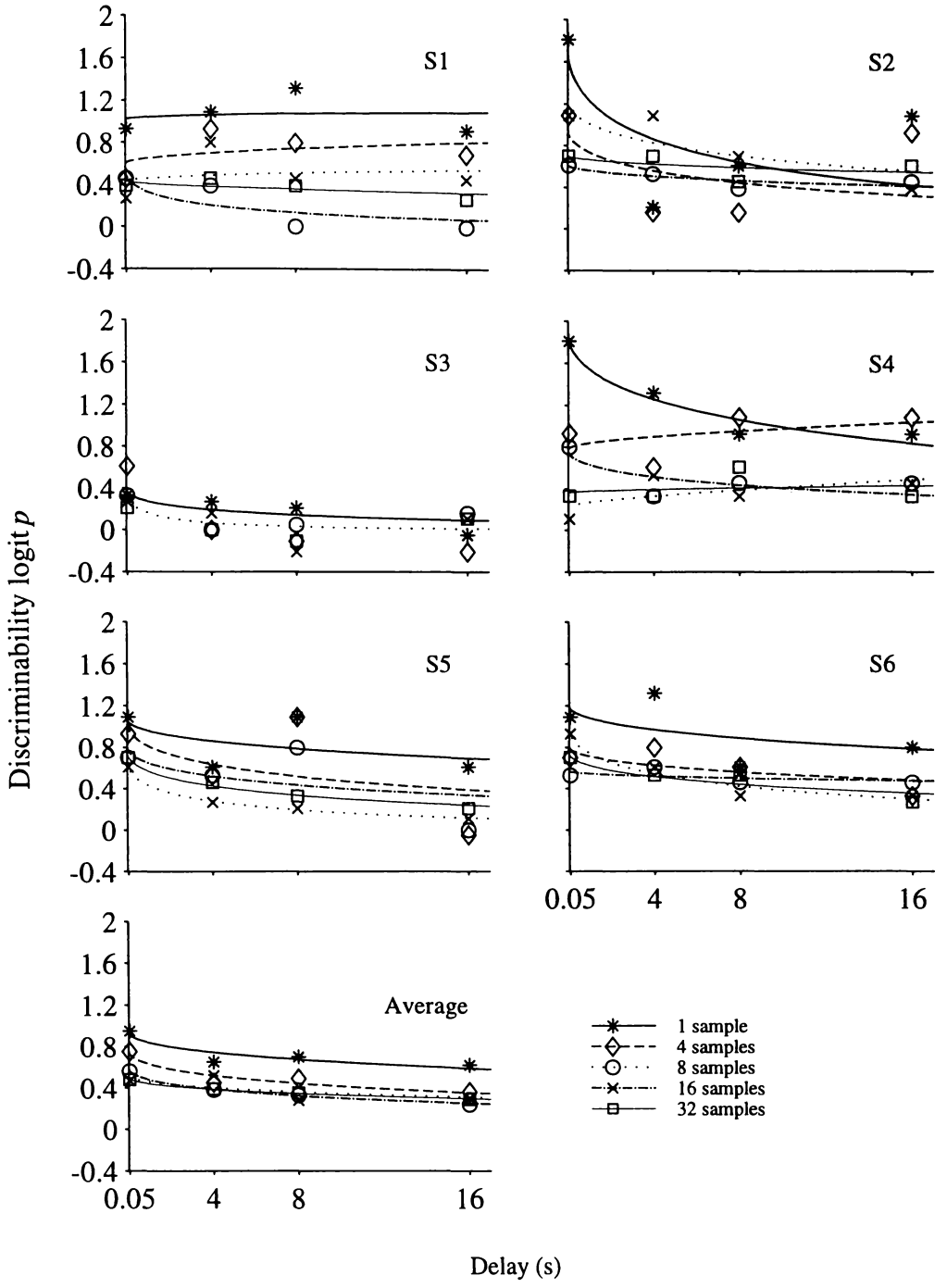

Figure 5.2. Logit *p* plotted as a function of delay and sample-set size for each participant and the average of these. The fitted functions are negative exponentials (Equation 2).

variance. It was found that this was not significant for delay ($W = 0.57$) but was for sample-set size ($W = 0.00$) and for the interaction ($W = 0.00$) ($\alpha < 0.05$). Given this the Huynh-Feldt (1970) correction was used when testing for significance in the ANOVA. For the ANOVA both the main effects of number of sample stimuli ($F = 10.06$, $df = 2.48, 12.42$) and delay ($F = 4.03$, $df = 3.00, 15.00$) were statistically significant ($\alpha < 0.05$) but the interaction ($F = 0.64$, $df = 10.91, 54.56$) was not significant.

Negative exponential functions (Equation 2) were fitted to the logit p data for each of the five conditions using non-linear regression. Table 5.3 shows the values obtained for the parameters a and b for both the individual and averaged data. The asterisk for S3 indicates where the function could not be fitted as the estimation process used to fit the function did not converge. Table 5.3 also shows the standard errors of estimate and the percentages of variance in the data accounted for (VAC) by the fitted functions. The number of sample stimuli had no systematic effect on b for the individual participants or for the averaged data. For S1 and S4 there were several negative b values which indicated that accuracy increased as the delay interval was increased.

For the individual participants a showed a general tendency to decrease as the number of sample stimuli was increased. For S1 a decreased systematically as the number of sample stimuli was increased and for S4 and S5 a decreased systematically with the exception of Condition 1, where there were 32 sample stimuli. For S6 a decreased as the number of sample stimuli was increased with the exception of Conditions 1 and 2, where there were 32 and 16 sample stimuli, respectively. Only for S2 and S3 was there no systematic tendency for a to decrease as the sample number was increased. For the averaged data a generally decreased as the number of sample stimuli was increased with the exception of Condition 3 where there were 8 sample stimuli. A unidirectional Ferguson's test for trend (Ferguson, 1965) showed that, across all participants, a trended significantly as the number of sample stimuli was increased ($\alpha < 0.05$). A unidirectional Ferguson's test also showed that b did not trend significantly as the sample-set size was increased ($\alpha < 0.05$). The data of S3 were not included in

Table 5.3 *Estimates of a, b, standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions (Equation 2) for each of the experimental conditions.*

	Sample N	*a*	*b*	VAC%	Std Err
S1	1	1.10	-0.02	4	0.20
	4	0.64	-0.10	22	0.16
	8	0.57	0.50	67	0.13
	16	0.45	-0.06	6	0.20
	32	0.46	0.07	33	0.06
S2	1	1.69	0.31	29	0.50
	4	1.06	0.29	14	0.43
	8	0.64	0.10	69	0.05
	16	1.34	0.20	69	0.18
	32	0.74	0.10	30	0.09
S3	1	0.41	0.35	61	0.10
	4	*	*	*	*
	8	0.41	0.88	58	0.08
	16	0.36	1.05	40	0.14
	32	*	*	*	*
S4	1	1.90	0.18	89	0.12
	4	0.82	-0.08	21	0.20
	8	0.81	0.21	58	0.12
	16	0.23	-0.20	42	0.13
	32	0.37	-0.05	5	0.12

S5	1	1.16	0.11	29	0.23
	4	1.02	0.24	30	0.40
	8	0.84	0.22	33	0.27
	16	0.70	0.44	100	0.04
	32	0.79	0.29	99	0.02
S6	1	1.32	0.11	27	0.28
	4	0.85	0.13	47	0.14
	8	0.59	0.05	28	0.06
	16	1.06	0.31	94	0.06
	32	0.79	0.19	84	0.07
Average	1	1.00	0.12	83	0.06
	4	0.80	0.20	90	0.05
	8	0.61	0.21	100	0.01
	16	0.54	0.13	50	0.08
	32	0.52	0.13	99	0.00

* Indicates that the negative exponential function could not be fitted to the data for these sample-set sizes.

these analyses as *a* and *b* values could only be produced for two of the conditions for this participant.

An analysis of response bias for a particular alternative, based on the size of the comparison stimuli, was carried out for the individual participants as a function of sample-set size and delays. The response bias was calculated as the logarithm of the ratio of the total number of times the small alternative was selected to the total number of times the large alternative was selected. Figure 5.3 shows that for three of the six participants (S1, S4, S6) this bias was towards selecting the smaller of the two comparison stimuli (values > 0). For the remaining three participants (S2, S3, S5) this bias was towards selecting the larger of the comparison stimuli (values < 0). There was no systematic change in this bias as the number of samples stimuli or the delay interval was increased, for any of the individual participants. An analysis of the average bias for the group showed that, overall, there was no bias (values close to 0) to any particular sized alternative.

An analysis of response bias for a particular alternative, based on the side of the screen comparison stimuli appeared on was also carried out for all individuals for each sample-set size and delay. Figure 5.4 shows that there was no large bias for selecting either the left or the right alternative for any of the individual participants (values all around 0). An analysis of the bias averaged across all of the participants also showed that, overall, there was no bias towards selecting either the left or the right alternative.

The individual data were analysed for any systematic changed in matching accuracy as the experimental session progressed. As matching accuracy varied across the experimental conditions, as a function of sample-set size, each condition was analysed separately. Each condition was split into two blocks of 16 trials for each of the four delay intervals. Matching accuracy for the first block, the first half of each condition as a function of delay interval, was compared to the second block of trials, the second half of each condition as a function of delay interval. There was no systematic increase or decrease in matching accuracy across the first and second blocks within each condition.

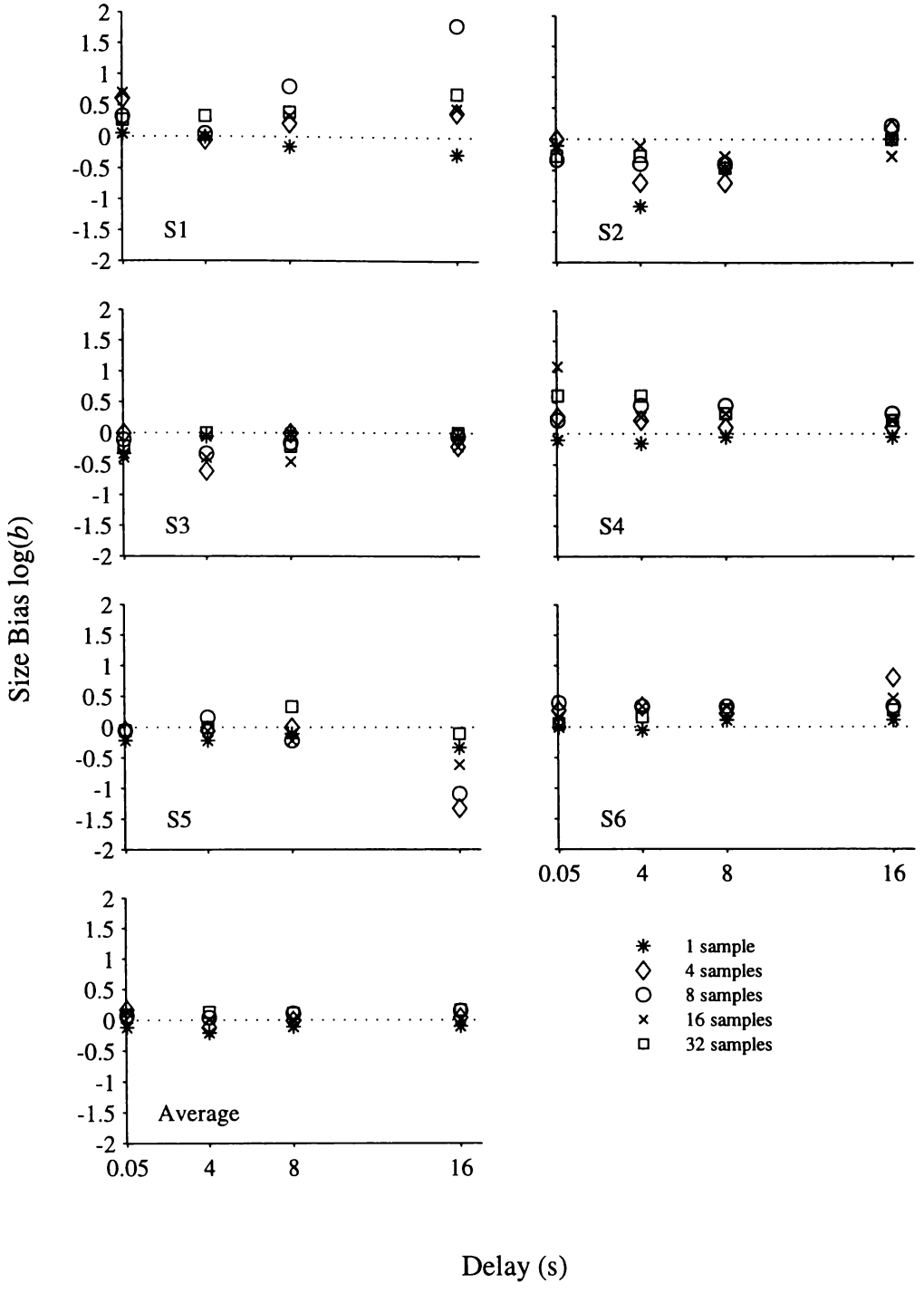

Figure 5.3. Log response bias plotted as a function of the size of the comparison stimuli for each participant and the average of these.

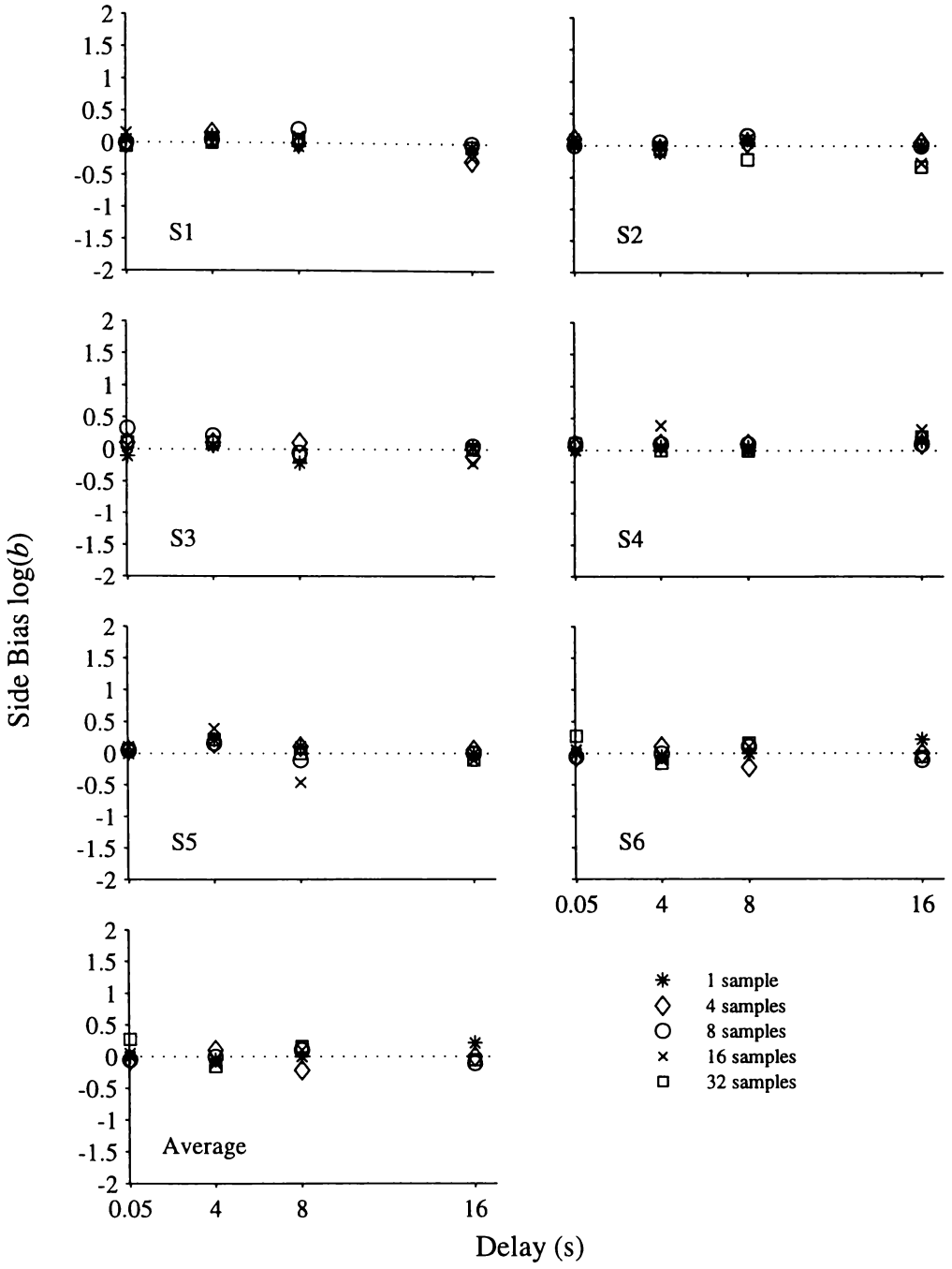

Figure 5.4. Log response bias plotted as a function of the side of the screen on which the comparison stimuli appeared, for each participant and the average of these.

Discussion

In the present study, accuracy decreased with increases in delay as shown by the *b* values for the fitted exponential functions. These *b* values were generally small and did not change systematically with changes in the number of sample stimuli. Accuracy generally decreased as the number of sample stimuli was increased as shown by the changes in *a*, initial discriminability. Thus, the sample-set size had an impact on initial discriminability, *a* not on the rate at which discriminability decreased with delay, *b*. These findings support the suggestion by White (1985) that *a* and *b* are independent.

For the average data initial discriminability, *a*, decreased most with the change from one to four samples, and decreased less than this with the change from four to eight samples. The changes in initial discriminability were smallest as the number of sample stimuli was increased from 8 to 16 sample stimuli and from 16 to 32 sample stimuli. Thus, had only one, four and eight sample stimuli been used the conclusion would have been that accuracy decreased as the sample-set size was increased. However, if only the 8, 16 and 32 sample-sets had been studied, the results would have suggested that sample-set size does not effect accuracy. These average data suggest that the result obtained when the number of sample stimuli is varied may depend on the sizes of the sample-sets which are used. For the individual data, if the sample-set size had been increased only from one to four to eight there would have been a systematic decrease in *a* as the sample-set size was increased for all of the participants with the exception of S2. If the sample-set size had been increased only from 8 to 16 to 32 then only S1 would have shown a systematic decrease in matching accuracy as the sample-set size was increased. Thus, the individual data also suggest that the result obtained when the number of sample stimuli are varied may depend on the sizes of the sample sets which are used.

Roberts' (1980) result, that accuracy decreased as the number of sample stimuli increased from one to two, is similar to that in the present study where a similar decrease was found as the number of sample stimuli increased. Additionally the average results for the 8, 16 and 32 samples here are similar to

the findings of Etkin and D'Amato (1969) and Mason and Wilson (1974), that there was no systematic change in accuracy over sample-set sizes from two to six. However, the present finding that accuracy tended to decrease systematically as the number of sample stimuli was increased from one to eight is not consistent with the findings of Etkin and D'Amato (1969) and Mason and Wilson (1974), for a similar range of sample-set sizes.

The present finding is not consistent with those obtained by Worsham (1975) and Mishkin and Delacour (1975), whose data suggest that matching accuracy increased as the sample-set size was increased. Inspection of Worsham's data show a clear increase in percent correct, over two, four and seven samples, only at the longer of the two delays studied. His finding of no change in percent correct, at the shorter of the two delays, is consistent with the results of Etkin and D'Amato (1969), Mason and Wilson (1974) and with those obtained for the larger sample sets for the average data in the present study. There is, however, no sign in the present data of an increase in accuracy as the sample-set size was increased, at any delay. It is not clear why the findings of these studies differ from the findings of the present study. It is however, clear that when studying the effect of sample-set on matching accuracy a wide range of sample-set sizes should be investigated.

The present data clearly show that individual participants had a tendency to select either the smaller or the larger of the two comparison stimuli. The fact that the participants show a systematic tendency to respond to either the larger or the smaller of the two comparison stimuli presents a potential problem with the use of logit p as a measure of discriminability. As White (1985) points out logit p only provides a good estimate of log d when there is little or not bias, which is not true of the individual data for the present study. However, he also suggests that it is relatively safe to assume that when data are averaged across participants there will be little or no bias. This suggestion relies on the assumption that if response bias is random then a similar number of participants will have biases to one or the other of the two alternatives and as such, the average response bias will be close to zero. The present study provides support for this suggestion in that three of the participants had biases towards the smaller and three of the participants had biases

towards the larger of the two comparison stimuli. Thus, the response bias seen, here, with the individual data was not apparent when the data were averaged across the participants. This suggests that, in the present study, only the average logit *p* data can be considered to provide a bias free measure of discriminability which is equivalent to log *d*.

It has been suggested that changes in matching accuracy as the number of sample stimuli is increased may be due to proactive interference. Edhouse and White (1988a) pointed out that when a subject responds incorrectly on a trial in a DMTS task, then behaviour is not under the control of the current sample stimulus. The response made may have been influenced by events which have occurred prior to the current trial. Such control by prior stimuli is generally termed proactive interference. There are a number of ways in which prior trials may influence accuracy on any given trial and, as a result, there appears to be no singular source of proactive interference. A number of different suggestions have been made about how proactive interference might effect matching accuracy in a DMTS task.

Etkin and D'Amato (1969) reasoned that, if proactive interference had a impact on matching performance in DMTS tasks, accuracy should be worse when the sample-set size was large. However, they do not state clearly why proactive interference would result in lower levels of accuracy when the number of sample stimuli used in a DMTS task was large (in the case of their study four was the largest sample-set size). Mason and Wilson (1974) pointed out that if the number of stimuli used as samples, on a fixed number of experimental trials, was increased then the number of times that a stimulus appeared as a sample would decrease. They suggested that this would increase the temporal discriminability of the sample stimuli and hence increase discriminability. Worsham (1975) and Roberts (1980) both suggested that proactive interference may also occur when the incorrect comparison stimulus on the present trial has appeared as the sample stimulus on the previous trial. In contrast to Etkin and D'Amato (1969), Worsham (1975) suggested that, if proactive interference was a significant factor in decreasing matching accuracy, matching should get better as the number of different sample stimuli was increased. He argued that the larger the number, the

less likely it is that the incorrect comparison stimulus will have appeared as the sample stimulus on the previous trial thus reducing this source of proactive interference. He suggested that as the number of sample stimuli was increased confusion resulting from the temporal recency of the stimuli was decreased.

Investigations of proactive interference in DMTS have focused on a number of variables including whether or not the sample stimuli on consecutive trials are the same. This is one aspect of what has been termed intertrial correspondence (ITC) (Edhouse & White, 1988a). Grant (1975) examined the effect of sample stimulus correspondence on matching accuracy using a DMTS procedure with pigeons. He found that when the sample stimuli on two consecutive trials were dissimilar, that is, did not correspond, matching accuracy was lower than when consecutive trials were the same. A similar finding was reported by Moise (1976) using monkeys as subjects on a DMTS task. Edhouse and White (1988a) also investigated the effect of ITC on matching accuracy using a DMTS procedure with pigeons. They found that while sample stimulus correspondence affected *b*, with a smaller *b* from trials with corresponding sample stimuli than from trials with non-corresponding sample stimuli, it had no apparent impact on *a*. Edhouse and White (1988b) used a DMTS task, with pigeons, to investigate the build up of proactive interference from sample-stimulus correspondence across the experimental session. They found that there was no evidence of a general performance decrement as the session progressed. These studies suggest that varying the probability of ITC in a DMTS task has an impact on matching accuracy, but that the effect is not cumulative.

DMTS data are typically analysed without taking the ITC, of sample stimuli, into consideration. This means that the typical DMTS result is actually a product of performance on trial pairs which correspond and trial pairs which do not. Increasing the number of different sample stimuli, when samples are presented randomly, will decrease the probability that the sample stimuli on any two consecutive trials will be the same. In this way, varying the sample-set size changes the probability of ITC of sample stimuli. Thus, given Edhouse and White's (1988a) finding, that ITC effected matching accuracy, the number of

sample stimuli used in a DMTS task might also be expected to have an affect on matching accuracy.

Given that the level of ITC may have varied as the number of sample stimuli were varied the data from the present study were reanalysed in order to investigate the effect of proactive interference from prior sample stimuli, or ITC. In order to examine ITC consecutive trials (*n* and *n*-1) were categorised into two types, corresponding trials where the sample stimulus was the same for both trials *n* and *n*-1 and non-corresponding trials where the sample stimulus was different for trials *n* and *n*-1. The data were analysed to examine whether the probability that the sample stimuli were the same on two consecutive trials changed as the number of sample stimuli was increased. The number of times that the sample stimulus on trial *n* and trial *n-1* were the same was counted for each participant for each sample-set size. The experimental probability of ITC for each of the conditions was then calculated by dividing the number of times that consecutive trials corresponded by the total number of stimuli in the sample-set. These empirical probabilities can be seen in Table 5.4 along with the theoretical probabilities for each of the conditions. The theoretical probabilities were calculated by dividing the number of times that each of the stimuli in each sample-set was repeated during a condition by the total number of stimuli in each set. Table 5.4 shows that as the number of stimuli in the sample-set was increased the overall probability of ITC decreased.

The data were then analysed in a manner similar to that used by Edhouse and White (1988a). For each participant the total number of consecutive trials on which the sample stimuli were the same was counted for each sample-set and each delay. The number of correct responses for both the corresponding and the non-corresponding trials were also calculated for all of the participants data. Logit *p* was obtained from these data for each type of trial pair, these are shown in Figure 5.5. For the individual data logit *p* was generally higher, across the delays, when trials *n* and *n* - 1 corresponded than when trials *n* and *n* - *1* did not correspond. Logit *p* for the average data was higher at all delays when trial pairs corresponded than when they did not correspond.

Table 5.4 *The number of sample stimuli in each condition, the number of times each stimulus was repeated and the theoretical and experimental probabilities that trials n and n-1 would correspond in each of the conditions*

Condition	Sample Number	Number of times each Sample stimulus was repeated	Theoretical probabilities (CT)	Experimental probabilities (CT)
1	1	32	1	0.975
2	4	8	0.250	0.259
3	8	4	0.125	0.137
4	16	2	0.063	0.069
5	32	1	0.000	0.001

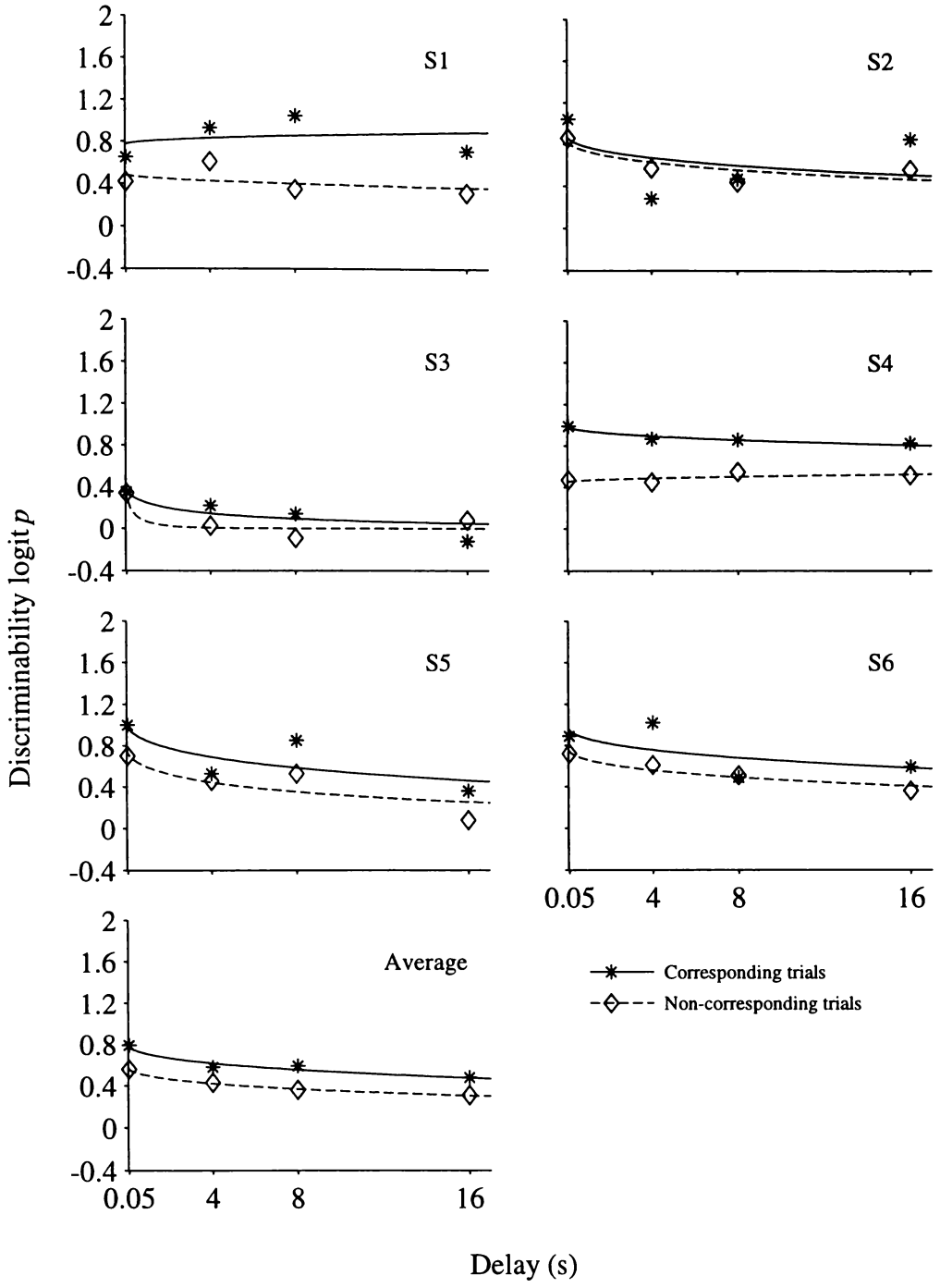

Figure 5.5. Logit *p* plotted as a function of delay and intertrial correspondence for each participant and the average of these. The fitted functions are negative exponentials (Equation 2).

Negative exponential functions (Equation 2) were fitted to both the individual and the average data. The values obtained for the parameters *a* and *b* are given in Table 5.5 along with the standard error of estimate and the percentage of variance in the data accounted for (VAC) by the fitted functions. For all participants, except S3, *a* was lower for non-corresponding trial types than for corresponding trial types. However, whether the sample stimuli on consecutive trials were the same or different did not have an impact on *b* for the individual data. For the average data *a* was lower for non-corresponding trial types than for corresponding trial types, and *b* was slightly larger for non-corresponding than for corresponding trials.

Another source of proactive interference may result from the non-matching comparison stimulus on trial *n* having appeared as the sample stimulus on trial *n-1*. The total number of times that the non-matching comparison stimulus on the present trial appeared as a sample stimulus on the prior trial, here, depended on the physical disparity used and ranged from 0.16% to 1.72% across the six participants. The non-matching comparison stimulus on trial *n* never appeared as a sample stimulus on trial *n-1* during the one or the four sample conditions, for any of the participants. The non-matching comparison stimulus on trial *n* never appeared as a sample stimulus on trial *n-1*, for S1 in the 32 sample condition and S2 in the 16 and 32 sample conditions. For trials where the non-matching comparison stimulus on trial *n* had been the sample stimulus on trial *n-1*, it was noted whether the participants were correct or incorrect on trial *n*. The percentages of correct responses on trial *n* ranged from 83% to 45% across the six participants. Thus, all but one of the participants (S3 (45%), the exception) were more likely to be correct on trial *n* than they were to be incorrect when the comparison stimulus on trial *n* had been the sample stimulus on trial *n*-1. For each of the participants matching accuracy was calculated for each of the sample-set sizes at each of the delays, excluding the data from trials where the non-matching comparison stimulus had been the sample stimulus on the previous trial. These values were then compared to matching accuracy for each of the sample-set sizes at each of the delay intervals where the non-matching comparison stimulus on the present trial had been the sample stimulus on the previous trial were included.

Table 5.5 *Estimates of log d_0, b, standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions for corresponding and non-corresponding trial types.*

	Trial Type	*a*	*b*	VAC %	Std Err
S1	CT	0.77	-0.04	9	0.15
	NCT	0.49	0.07	15	0.10
S2	CT	0.89	0.14	12	0.28
	NCT	0.85	0.15	71	0.08
S3	CT	0.42	0.53	68	0.10
	NCT	0.55	2.12	83	0.20
S4	CT	0.99	0.05	92	0.02
	NCT	0.45	-0.04	40	0.03
S5	CT	1.03	0.20	58	0.17
	NCT	0.78	0.28	69	0.13
S6	CT	0.98	0.13	41	0.17
	NCT	0.77	0.16	93	0.04
Average	CT	0.80	0.13	95	0.02
	NCT	0.58	0.16	100	0.01

There were no systematic increases or decreases in matching accuracy when these trial were excluded in comparison to when they were included. Any differences in accuracy were small.

In this study matching accuracy was higher across all delays when trial pairs corresponded than when they did not (Figure 5.5). Given that *b* is very similar for the two functions these data do not provide support for Edhouse and White's (1988a) conclusion that ITC affects *b* not *a*. It is not clear why this is so. The present result does, however, suggest that because performance on non-corresponding trial pairs is lower than performance on corresponding trial pairs then, if the number of non-corresponding trials is increased, the overall measure of accuracy would decrease. Given the fixed number of trials in any condition and the increase in the number of stimuli that appear as samples, the occurrence of corresponding trial pairs decreased as the sample-set size was increased. In fact, when there were 32 sample stimuli each trial pair was unique. Thus there is a possible confound here, in that as the sample-set size is increased the frequency of non-corresponding trials is also increased.

Thus, it is not clear whether the level of ITC of sample stimuli or the sample-set size gave rise to the apparent decrease in performance which occurred as the sample-set size was increased. To clarify this it is necessary to vary the probability of corresponding and non-corresponding trial pairs independently of the number of sample stimuli. This could be done by keeping the sample-set size constant and deliberately varying the number of corresponding and non-corresponding trial pairs in a manner similar to that of Edhouse and White (1988a).

As the present study found evidence of proactive interference from sample stimuli between consecutive trials it might be argued that this interference would build up across the experimental session in a cumulative manner. Edhouse and White (1988b) investigated this suggestion for their data and found no evidence that interference from prior trials built up as the experimental session progressed. The present study supports Edhouse and White's (1988b) finding in that no evidence of a cumulative build up of proactive interference, from sample stimuli was found.

Previous studies (Worsham, 1975 and Roberts, 1980) have suggested that the incorrect comparison stimulus on the present trial having appeared as the sample stimulus on the preceeding trial may also be a source of proactive interference and hence may influence matching accuracy. In this study it was not intended that there be any overlap between the comparison stimuli which were generated from the sample stimuli. However, depending on the result of the titration procedure, the number of stimuli required meant that it was possible for some overlap between the sample and comparison stimuli to occur, particularly with more than eight sample stimuli. Because of this, it was possible for the incorrect comparison stimulus on the present trial to have appeared as the sample stimulus on prior trial. This did not occur very often, the frequency with which it did varied across the participants and across the various sample-set sizes. These types of trial pairs never occurred during the one or four sample conditions for any of the participants.

If Worsham (1975) and Roberts (1980) are correct it could be argued that the decrease in accuracy which occurred as the number of sample stimuli was increased might be partially accounted for by an increase in the number of times that the comparison stimulus had appeared as a sample stimulus on the prior trial. However, even when these types of trials did occur they did not have an impact on accuracy. Analysis showed that performance was more likely to be correct than to be incorrect on such trials. Given this, a comparison stimulus on the current trial appearing as the sample stimulus on the prior trial could not account for the decrease in accuracy that occurred as the number of sample stimuli was increased. As there was no evidence of proactive interference occurring as a result of the incorrect comparison stimulus on the present trial having appeared as the sample stimulus on the prior trial, there was no apparent need to investigate for any cumulative effect across the experimental session.

There is a potential explanation for the difference between the results of this study and those of Worsham (1975) and Mishkin and Delacour (1975), whose data suggest that matching accuracy increased as the sample-set size was increased. As shown above, in the present study, the incorrect comparison stimuli on the present trial did not appear as the sample stimuli on the prior trial very

often. In both Worsham's (1975) and Mishkin and Delacour's (1975) studies the sample stimuli were also the comparison stimuli, which was not generally true for the present study. This means that in their respective studies incorrect comparison stimuli are likely to have appeared as sample stimuli on the prior trial on a more frequent basis than in the present study. If this were the predominant source of proactive interference, as opposed to proactive interference from sample stimuli, then as the sample-set size is increased it is less likely that an incorrect comparison stimulus will have appeared as a sample stimulus on the prior trial. If this type of proactive interference acted in much the same way as that which is caused by sample stimulus correspondence then you might expect that the incorrect comparison stimulus having appeared as the sample stimulus on the prior trial would result in lower performance than when the incorrect comparison was not the previous sample stimulus. If the likelihood that the present incorrect comparison stimulus had appeared as the sample stimulus decreased as the number of sample stimuli was increased you might then expect accuracy to also increase. Thus, the result that they obtained might be expected.

In summary the present data set shows that sample stimulus number affects *a* with human performance on a DMTS task, but does not appear to affect *b*. However, it is not clear whether this effect is a product of the increase in the sample-set size or of the associated decrease in the correspondence of sample stimuli on consecutive trial pairs.

Experiment 6

The previous experiment found that as the number of sample stimuli used in a DMTS task was increased matching accuracy decreased. This result was consistent with those of Roberts (1980) and Adamson (1995). It was suggested that increasing the number of sample stimuli affected *a*, initial discriminability, but did not effect *b*, rate of decay. The decrease in *a* when the number of sample stimuli was increased from one to four sample stimuli was larger than the decrease in *a* when the number of sample stimuli was increased from four to eight. It was suggested that the decrease in the size of the changes in *a,* as the number of sample stimuli was increased, might be due to a corresponding decrease in ITC, i.e., to the probability that any two consecutive trials will have the same sample stimulus.

It was suggested that this could be investigated further by changing the probability of ITC without changing the number of sample stimuli. The probability of ITC could be changed by organising the sequence in which the sample stimuli appeared to give different ratios of corresponding and non-corresponding trials, across a set number of trials. This would require at least two sample stimuli.

The previous experiment did not include a two-sample condition, and it was not known how matching accuracy with two sample stimuli would compare to matching accuracy with other sample-set sizes. Given the results of the previous experiment, it would be expected that matching accuracy with two samples would be lower than with one sample but higher than with four samples. However, as the magnitude of change in *a* decreased as the number of sample stimuli was increased, for the previous experiment, it is not clear exactly what *a* value would be expected.

The aim of the present experiment was to replicate the previous experiment systematically, with the addition of a two-sample condition. The question was, whether or not the magnitude of the change in *a* between the one and the two-sample sets and between the two and the four-sample sets would be the same. The 32 sample condition was not included in the present experiment as

there were no systematic differences between the *a* values obtained for the 16 and the 32 sample conditions in the previous experiment.

Method

Participants

Five first year psychology students participated in this experiment. Each participant received course credit towards a first year psychology courses, irrespective of their performance on the experimental task.

Apparatus

The apparatus used was the same as that used in Experiment 1.

Stimuli

Filled white disks of varying sizes, presented on a blue background, were used as stimuli. The sample stimuli were drawn from a pool of 32 disks ranging in size from 66 pixels to 111 pixels in diameter, in steps of 3 pixels. There were 5 different sized sample sets with 1, 2, 4, 8 or 16 stimuli. When there was one sample stimulus it was 87 pixels in diameter. The 2, 4, 8 and 16 sample-stimulus sets contained, respectively, all the stimuli between and inclusive of 84 to 93, 78 to 99 and 66 to 111 pixels in diameter as shown in Table 6.1. The non-matching comparison stimuli differed in size from the sample stimuli by a proportion of the diameter of the sample stimulus. For each sample stimulus there were two possible non-matching sample stimuli, one larger and one smaller than the sample stimulus. The exact proportional difference for each participant was selected during the initial part of the experiment.

Procedure

The general procedure was similar to that used in Experiment 5. During the initial part of the experimental session a one-sample titration procedure was used to select the proportional difference between the sample stimuli and the non-matching comparison stimuli. The titration procedure ended when participants reached a criterion of 100% correct across 30 trials. The main part of

Table 6.1 *The size of the stimuli pixels in diameter, in the sample pool and the sample-set size where the sample appeared as a stimulus.*

Sample size	Sample-set size where the sample appeared as a stimulus
66	16
69	16
72	16
75	16
78	8, 16
81	8, 16
84	4, 8, 16
87	1, 2, 4, 8, 16
90	2, 4, 8, 16
93	4, 8, 16
96	8, 16
99	8, 16
102	16
105	16
108	16
111	16

the session involved a DMTS task. This experiment differed from Experiment 5 in that there were 16 sample stimuli in Condition 1, 8 sample stimuli in Condition 2, 4 sample stimuli in Condition 3, 2 sample stimuli in Condition 4 and 1 sample stimulus in Condition 5. The delay intervals were blocked in the following order; 0.05, 4, 8 and 16 s and there were 32 trials for each of the conditions at each of the delay intervals. There was a total of 160 trials in each condition across all four delay intervals. The experimental session ended when the participant had completed 800 trials at the end of Condition 5. Once the experimental session had ended the participants were given information about the purpose of the experiment.

Results

S1 to S5 completed the titration in 80, 60, 150, 70 and 80 trials and reached proportional disparities of 14, 10, 30, 12 and 10%, respectively. S3 did not meet the pre-set criterion level of accuracy (S3 obtained 97% correct) at the maximum possible physical disparity (30%), thus, they started the DMTS with this maximum physical disparity. S2 and S5 started the DMTS task with the smallest physical disparity relative to the other participants. The proportional disparities here are similar to those in the previous experiment.

Figure 6.1 shows percent correct for each of the five participants, and the averaged data, for each of the sample-set sizes as a function of delay. For the individual data, and hence also the average data, accuracy tended to decrease as the delay interval was increased and accuracy also tended to decrease as the number of sample stimuli was increased. Matching accuracy, for the average data, was highest, for the 0.05 s delay, when there was only one sample stimulus. When there was only one sample stimulus matching accuracy for the average data was highest at the 0.05 s delay. When there were two sample stimuli matching accuracy was higher at 0.05 s than it was at this delay when there were four and eight sample stimuli. Matching accuracy when there were 8 and 16 sample stimuli was similar at the 0.05 s delay.

Figure 6.2 shows logit p, calculated using Equation 3, for each of the five participants, and the average of these, as a function of delay for each experimental

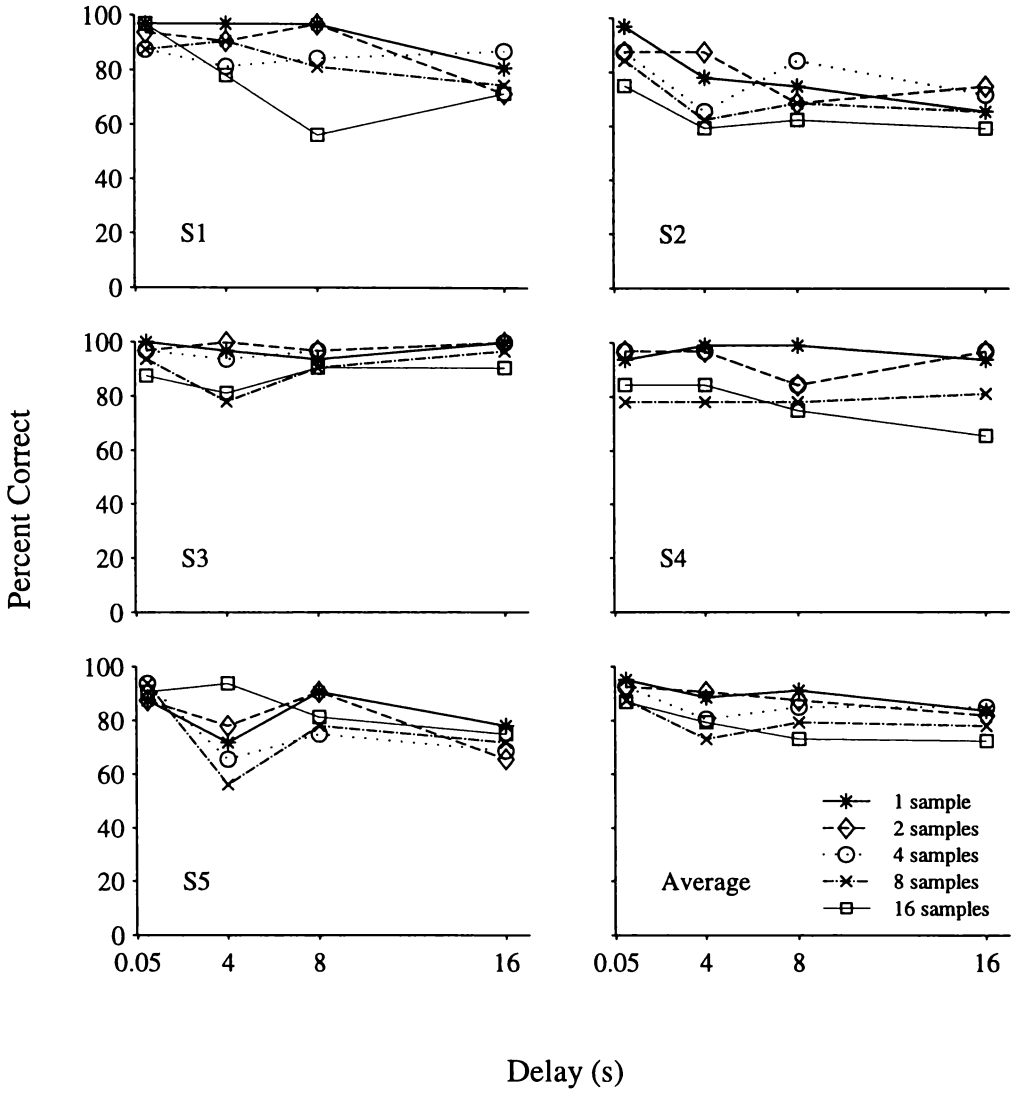

Figure 6.1. Percent correct as a function of delay and sample-set size for each participant and the average of these.

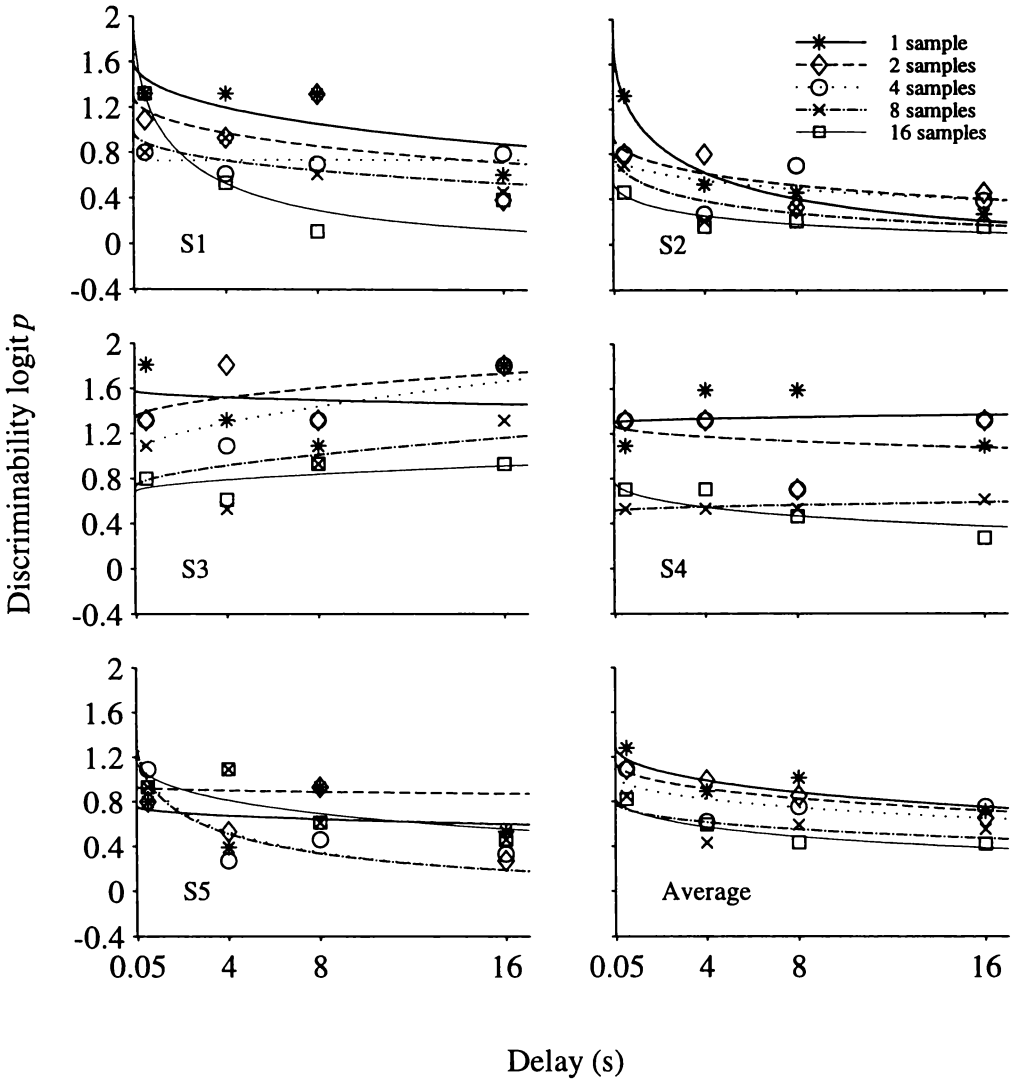

Figure 6.2. Logit *p* plotted as a function of delay and sample-set size for each participant and the average of these. The fitted functions are negative exponentials (Equation 2).

condition. For S3 there were five cases where there were zero errors but there were no cases of zero error for any of the other four participants. Thus, as for the previous experiments, the Hautus (1995) correction was used. The same trends are evident in the logit *p* data as were evident in the percent correct data. A repeated measures ANOVA was carried out on these data. Prior to this Mauchly's test of sphericity was used to test the data for homogeneity of variance. This test was not significant for delay (W = 0.12) or sample-set size (W = 0.17) but was for the interaction (W = 0.00) (α = 0 05). Thus, the Huynh-Feldt (1970) correction to the df was used here. For the ANOVA the main effect of sample-set size (F = 7.06, df = 1.88, 7.50) was significant but the main effect of delay (F = 2.44, df = 2.69, 10.76) and the interaction (F = 0.83, df = 12.00, 48.00) were not significant ($\alpha > 0.05$).

Negative exponential functions (Equation 2) were fitted to the logit *p* data for each of the five conditions using non-linear regression. Table 6.2 shows the values obtained for the parameters *a* and *b* for both the individual and averaged data as well as the standard errors of estimate and the percentages of variance in the data accounted for by the fitted functions (VAC). There was a systematic decrease in *a* as the number of sample stimuli was increased for S3. For the remaining participants *a* generally tended to decrease as the number of sample stimuli was increased with the exception of one or two of the conditions. For the averaged data *a* decreased as the number of sample stimuli was increased, the exception being from 8 to 16 sample stimuli. A unidirectional Ferguson's (1965) test showed that *a* trended significantly as the number of sample stimuli was increased ($\alpha < 0.05$). A unidirectional Ferguson's test also showed that *b* did not trend significantly as the number of sample stimuli was increased for the individual or the averaged data ($\alpha < 0.05$). In the case of S1, S3 and S4 the negative *b* values indicate that accuracy increased as the delay interval was increased.

In most cases the percentage of VAC was low for all fitted functions, with the exception of some conditions, but the percentage of VAC did not vary systematically with experimental condition. For the averaged data the percentage of VAC was generally higher than that obtained for the individual data. The

Table 6.2 *Estimates of a, b, (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions for each of the experimental conditions for each participant and the average of these.*

	Samples	*a*	*b*	VAC%	Std Err
S1	1	1.62	0.15	53	0.21
	2	1.31	0.15	29	0.29
	4	0.72	-0.01	1	0.08
	8	0.99	0.15	58	0.12
	16	2.16	0.71	87	0.16
S2	1	1.91	0.55	97	0.07
	2	0.98	0.22	59	0.13
	4	0.81	0.18	26	0.19
	8	0.85	0.39	69	0.11
	16	0.58	0.41	70	0.06
S3	1	1.58	0.02	1	0.31
	2	1.33	-0.07	26	0.21
	4	1.01	-0.13	55	0.18
	8	0.72	-0.12	18	0.26
	16	0.68	-0.07	30	0.11
S4	1	1.30	-0.01	1	0.25
	2	1.28	0.04	6	0.26
	4	1.28	0.04	6	0.26
	8	0.51	-0.03	55	0.02
	16	0.79	0.19	70	0.10

S5	1	0.76	0.06	5	0.21
	2	0.93	0.17	29	0.21
	4	1.47	0.51	76	0.16
	8	1.41	0.50	50	0.25
	16	1.19	0.19	58	0.16
Average	1	1.30	0.14	84	0.08
	2	1.16	0.12	90	0.05
	4	1.04	0.12	52	0.12
	8	0.80	0.13	50	0.11
	16	0.85	0.20	96	0.03

standard error of estimate was generally small for all of the data sets and for the averaged data.

An analysis of response bias for a particular alternative, based on the size of the comparison stimuli, was carried out for the individual participants for each sample-set size and delay interval. As in Experiment 5 response bias was calculated as the logarithm of the ratio of the total number of times the small alternative was selected to the total number of times the large alternative was selected. Figure 6.3 shows that with the exception of S2, biases were close to zero and there were no consistent biases towards selecting either the larger or the smaller of the two comparison stimuli. For S2 there was a consistent bias towards selecting the smaller of the two comparison stimuli (values > 0). There was no systematic change in bias as the number of sample stimuli was increased from 1 to 16 for any participant. Figure 6.3 also shows that for the average data there were no consistent biases towards any particular sized comparison stimulus.

An analysis of response bias for a particular alternative, based on the side of the screen that the comparison stimuli appeared on, was also carried out for all individuals as a function of sample-set size and delay interval. Figure 6.4 shows that in all cases, including the average data, there was no consistent response bias for selecting either the left or the right alternative.

The individual data were also analysed for any systematic change in matching accuracy as the experimental session progressed. Data from each condition were analysed separately. Each condition was split into two blocks of 16 trials as a function of the delay. Matching accuracy for the first block of 16 trials was then compared to matching accuracy for the second block of 16 trials, as a function of delay interval. There was no systematic change in matching accuracy across the first and second blocks of trials for any of the experimental conditions and these data are not presented.

An analysis of ITC was carried out for both the individual data and the averaged data, as in Experiment 5. Logit p was obtained from these data for each type of trial pair, these are shown in Figure 6.5. For all of the participants with the exception of S5 logit p was higher at all delays when trial pairs corresponded than when they did not correspond. Logit p generally decreased as the delay

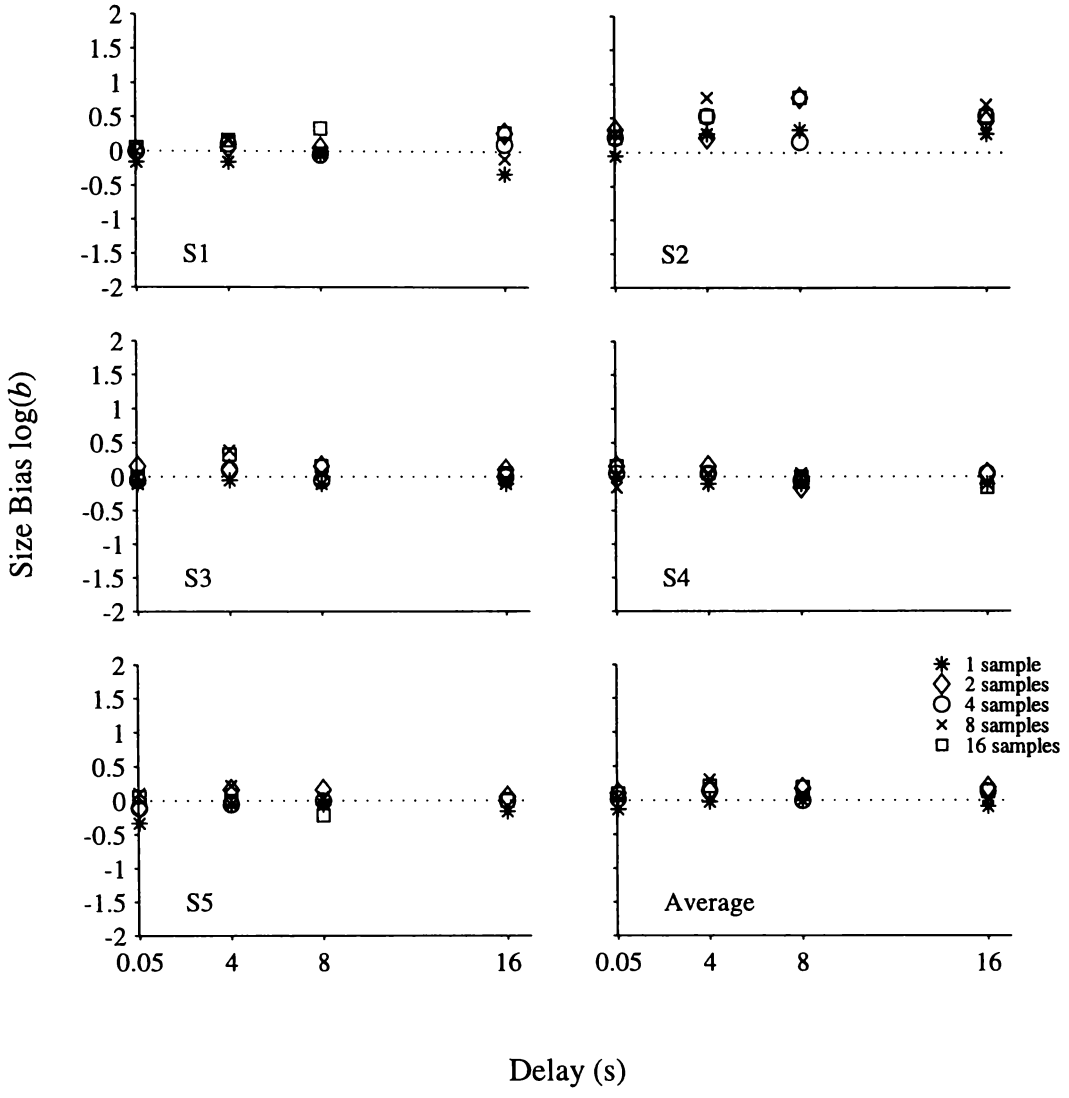

Figure 6.3. Log response bias plotted as a function of the size of the comparison stimuli for each participant and the average of these.

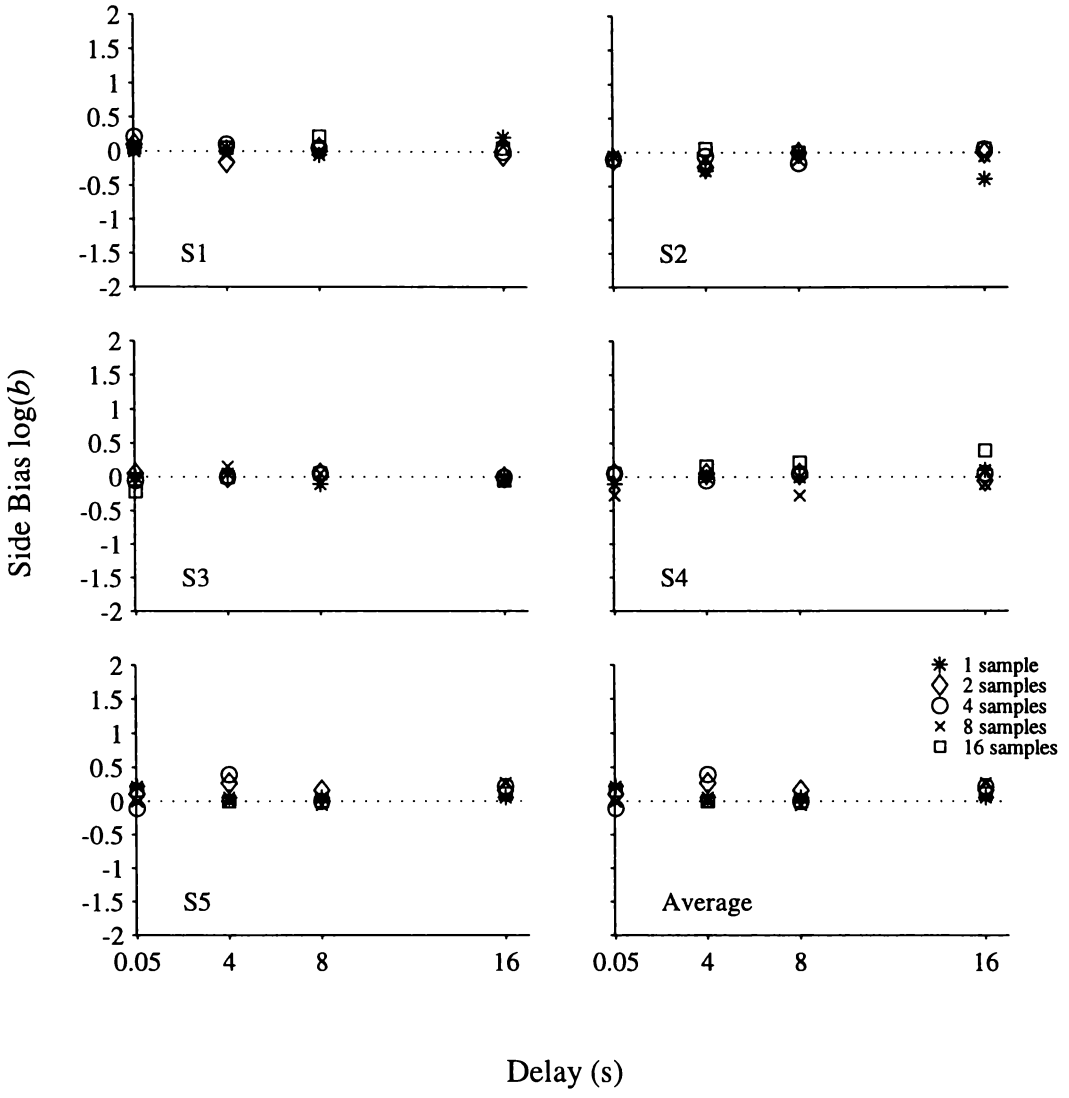

Figure 6.4. Log response bias plotted as a function of the side of the screen on which the comparison stimuli appeared, for each participant and and the average of these.

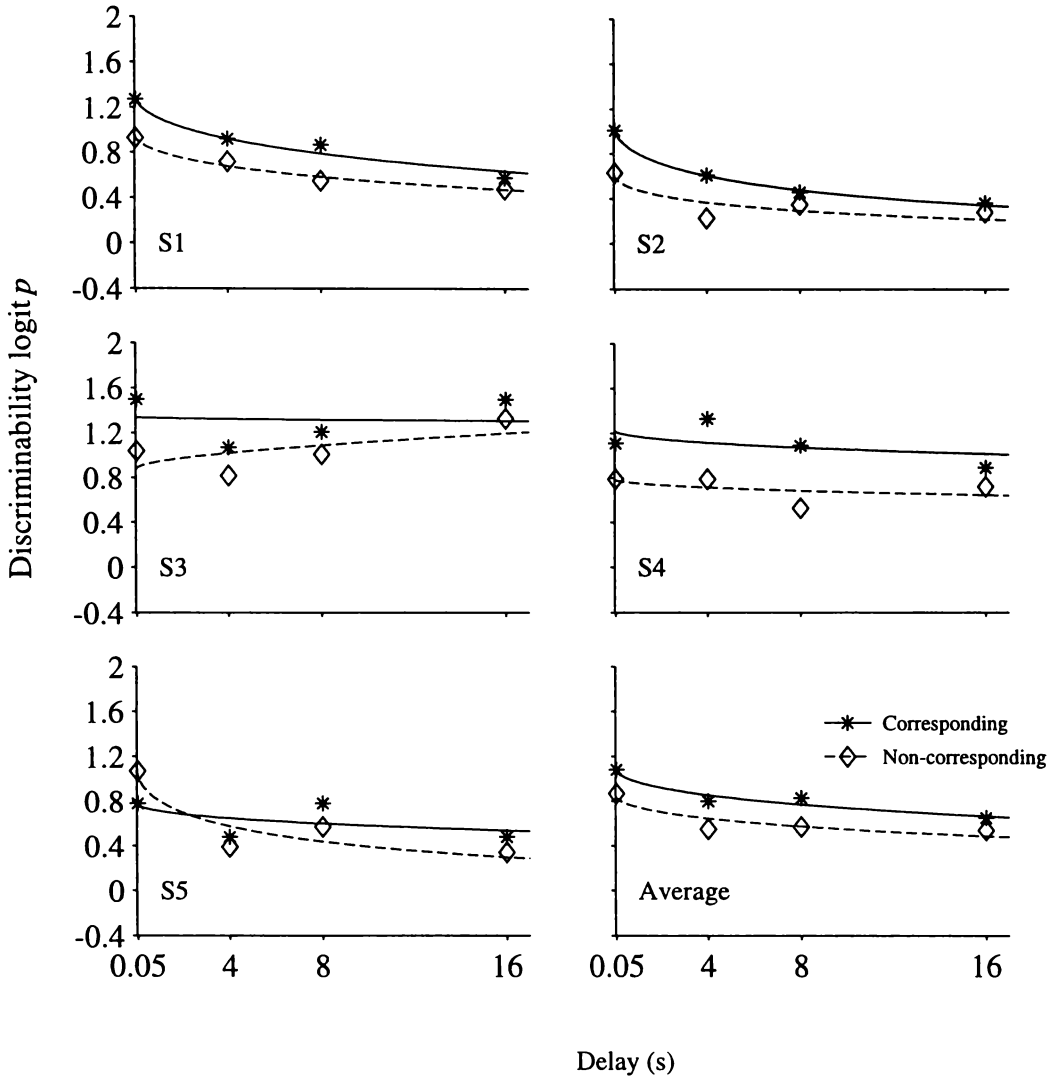

Figure 6.5. Logit *p* as a function of delay and intertrial correspondence for each of the participants and the average of these. The fitted functions are negative exponentials (Equation 2).

interval increased, irrespective of whether trials corresponded. For the average data logit *p* was higher at all delays when trial pairs corresponded than when they did not correspond.

Negative exponential functions (Equation 2) were fitted to these data. The values obtained for the parameters *a* and *b* are given in Table 6.3, along with the standard error of estimate and the percentage of VAC by the fitted functions. For all of the participants, except S5, *a* was lower for non-corresponding trial types than for corresponding trial types, but *b* was not effected by trial type. For the average data *a* was also lower for non-corresponding trial types than for corresponding trial types, but *b* was not effected by trial type. The percentage of VAC by the fitted function was high for S1, S2, and for the averaged data, and was low for S3 and S4 for both corresponding and non-corresponding trials. For S5 the percentage of VAC by the fitted function was low for corresponding trials and high for non-corresponding trials. The standard error of estimate was generally small for all of the participants and for the averaged data, for both corresponding and non-corresponding trials.

The individual data were also analysed in order to examine whether any decrease in accuracy had resulted from the non-matching comparison stimulus on trial *n* having appeared as the sample stimulus on trial *n*-1. The total number of times that the non-matching comparison stimulus on trial *n* had appeared as the sample stimulus on trial *n*-1 ranged from 0.00 to 1.25%, of total trials, across the five participants. For S3 and S5 the non-matching comparison stimulus on trial *n* never appeared as the sample stimulus on trial *n*-1 for any of the conditions. The non-matching comparison stimulus on trial *n* never appeared as a sample stimulus on trial *n*-1 during the one or the two sample condition, for S1, S2 or S4 or during the four sample condition for S4. For those trials where the non-matching comparison stimulus on trial *n* had been the sample stimulus on trial *n-1*, it was noted whether the participants were correct or incorrect on trial *n*. It was found that participants were more likely to be correct than incorrect on these trials with the percentage of correct responses on trial *n* ranging from 60% for S1 to 86% for S2 and S4.

Table 6.3 *Estimates of a, b (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions for corresponding and non-corresponding trial pairs n and n - 1 and the average of these.*

	Trial Type	*a*	*b*	VAC %	Std Err
S1	CT	1.33	0.18	96	0.05
	NCT	0.97	0.18	97	0.03
S2	CT	1.08	0.29	100	0.01
	NCT	0.63	0.27	73	0.08
S3	CT	1.34	0.01	1	0.19
	NCT	0.87	-0.08	33	0.15
S4	CT	1.23	0.05	25	0.14
	NCT	0.79	0.05	20	0.10
S5	CT	0.78	0.09	29	0.13
	NCT	1.11	0.33	84	0.12
Average	CT	1.10	0.13	94	0.04
	NCT	0.86	0.14	82	0.06

Discussion

The present experiment supports the finding of the previous experiment that increasing the sample-set size results in a decrease in accuracy at zero delay, *a*, but does not effect the rate at which accuracy decreases as the delay is increased, *b*. The present study also found that accuracy for the two-sample condition tended to fall between accuracy for the one- and the four-sample condition. Accuracy generally decreased as the number of sample stimuli was increased as shown by the *a* values for the fitted exponential functions. In addition, *b* did not change systematically as the number of sample stimuli was increased.

In the previous experiment it was suggested that the result obtained when the number of sample stimuli are varied may depend on the sample-set sizes used. The present result generally confirmed this. However, here, the results do not provide support for the finding of the previous experiment that the magnitude of the change in *a* which occurred as the sample set was increased changed.

The present data were analysed for any systematic tendency for the participants to select either the smaller or the larger of the two comparison stimuli. With the exception of S2 there was no systematic tendency for the participants to select either the larger or the smaller comparison stimulus. This finding differs from that of the previous experiment where all six participants showed a systematic tendency to choose either the larger or the smaller of the two comparison stimulus. This means that here logit *p*, for both the individual and the averaged data here, provides a good estimate of log *d*. Thus, the present experiment is a replication of the findings of the previous experiment, that matching accuracy decreases as the sample-set size is increased, which is not confounded by bias.

The present data showed that matching accuracy was higher across all delays when trial pairs corresponded than when they did not correspond, with the exception of S5 for a delay of 0.05 s. This is reflected in *a*, which was larger for corresponding trial pairs than for non-corresponding trial pairs for four out of the

five participants. Thus the results of the present study are consistent with the findings of the previous experiment that ITC effected *a* but did not effect *b*.

As noted previously it has been suggested that proactive interference may occur when the incorrect comparison stimulus on the present trial has appeared as the sample stimulus on the preceding trial. An analysis of these trials here showed that they did not have a large effect on accuracy when they did occur and that participants were more likely to be correct than incorrect for these types of trials. This also supports the previous suggestion that the decrease in accuracy which occurred as the number of sample stimuli was increased could not be attributed to the non-matching comparison stimulus having appeared as the sample stimulus on the prior trial.

In summary the results, here, support the finding of the previous experiment that increasing the number of sample stimuli effects *a* but does not effect *b*. The present result showed that *a* for the two sample condition tended to fall between *a* for the one and the four sample conditions and this change in *a* tended to be smaller from the one to the two sample condition than from the two to the four sample condition. The present result established where *a* for a two sample condition might be expected to fall with respect to other sample-set sizes. Given this, the intention of the next experiment is to vary the probability of ITC without varying the sample-set size.

Experiment 7

The previous two experiments found that matching accuracy decreased systematically as the number of sample stimuli used for the DMTS task was increased. In addition, as Edhouse and White (1988a) found, matching accuracy was lower on corresponding trials than on non-corresponding trials. However, for the previous experiments sample stimulus number and the number of corresponding trials co-varied and so it is not clear if it was the decrease in ITC or the increase in the number of sample stimuli that gave rise to these results. As suggested previously this could be investigated by holding the number of sample stimuli constant and changing the ratio of corresponding to non-corresponding trials. For example, if there were 32 trials and the probability of ITC was equal to 0.50 there would need to be 16 corresponding and 16 non-corresponding trials. If the probability of ITC was 0.25, there would need to be 8 corresponding and 24 non-corresponding trials. Thus, if the number of sample stimuli used was kept constant, the effect of ITC on matching accuracy could be examined independently.

This was done in the present experiment. The probabilities of ITC used here, (probabilities = 1.0, .5, .25, .125 and .0625) corresponded to the probabilities of ITC which occurred in the previous experiment for the sample set sizes 1, 2, 4, 8 and 16. The number of sample stimuli was kept at two except in the condition in which the probability of ITC was equal to 1, where only one-sample stimulus was used.

Method

Subjects

Four graduate level psychology students participated in this experiment. Each participant received a book voucher for their participation, irrespective of their performance on the experimental task.

Apparatus

The apparatus used was the same as that used in Experiment 1.

Stimuli

The stimuli were filled white disks 60 and 63 pixels in diameter, presented on a blue background. The non-matching comparison stimuli differed in size from the sample stimuli by a proportion of the diameter of the sample stimulus. For each sample stimulus there were two possible non-matching sample stimuli, one larger and one smaller than the sample stimulus.

Procedure

The general procedure was the same as that used for Experiment 1. The instructions for the present experiment differed from those given to the participants for Experiment 1 as described below. The participants were informed that sometimes during the experimental procedure the computer screen would turn red for a short period of time, and that a new trial would not start until the screen turned blue again. They were also informed that two breaks were programmed during the experimental session. They were instructed that breaks would be signalled by a message box on the screen which said click OK to continue. When this message appeared they could take a break, for as long as they wished, and could begin the experiment again when they were ready by clicking on the OK button.

The experimental session began with a titration similar to that used in Experiment 2 but here two sample stimuli were used. The two sample stimuli were 60 and 63 pixels in diameter. As in Experiment 2 there was a maximum of 200 trials in the titration but if the participant reached the pre-set criterion of 100% over a total of 30 trials the titration procedure ended. At the end of the titration task the DMTS task began.

The main part of the session involved a DMTS task. The general procedure was the same as used in Experiment 2 except that delays 0.05, 4, 8 or 16 s were used. There were six conditions in which the probability of ITC was varied. In Condition 1 the probability of ITC was set at 1.0 which meant that only the 63-pixel sample stimulus was used. In Condition 2 the probability of ITC was set at .0625, Condition 3 it was set at .5, Condition 4 it was set at .125, Condition 5 it was set at .25 and in Condition 6 it was set at 1.0 There were a

total of 160 trials in each of the conditions, with 32 trials at each of the delay intervals. For each condition the total number of corresponding trials required to give the level of probability was calculated (For example, 16 corresponding trials out of 32 gives a probability of 0.5). These corresponding trials were spaced evenly across the block of 32 trials for each of the delays. Each condition ended after the completion of 160 trials and the next condition began immediately. The experimental session ended when the participant had completed 960 trials at the conclusion of Condition 6. Once the experimental session had ended the participants were given Information about the purpose of the experiment.

Results

S1 to S4 completed the titration in 50, 40, 60 and 50 trials respectively. They reached proportional disparities of 9, 11, 9 and 11% respectively. Unfortunately it was later discovered that the titration procedure had not functioned correctly for S2 and S4 and the titration procedure ended before they were performing at the criterion level. For S2 the titration procedure finished when the participant had reached a proportional disparity of 7% and had obtained 83% correct across a total of 30 trials. For S4 the titration procedure finished when the participant had reached a proportional disparity of 9% and had obtained 97% correct across a total of 30 trials. When the DMTS task began the proportional disparity for both participants had stepped out to 11%.

Figure 7.1 shows percent correct and Figure 7.2 shows logit p, calculated using Equation 3, for each of the four participants and the average of these as a function of delay for each probability of ITC. Generally accuracy at the 0.05 s delay was high for all levels of ITC falling between 100 and 90% correct, with the exception of the ITC of 0.25 for S2. For S3 there were zero errors in five cases and for S4 there were zero errors in four cases. For the individual data accuracy decreased slightly as the delay was increased but there were no apparent systematic changes in accuracy as the probability of ITC was decreased. For the average data accuracy decreased slightly as the delay interval was increased and

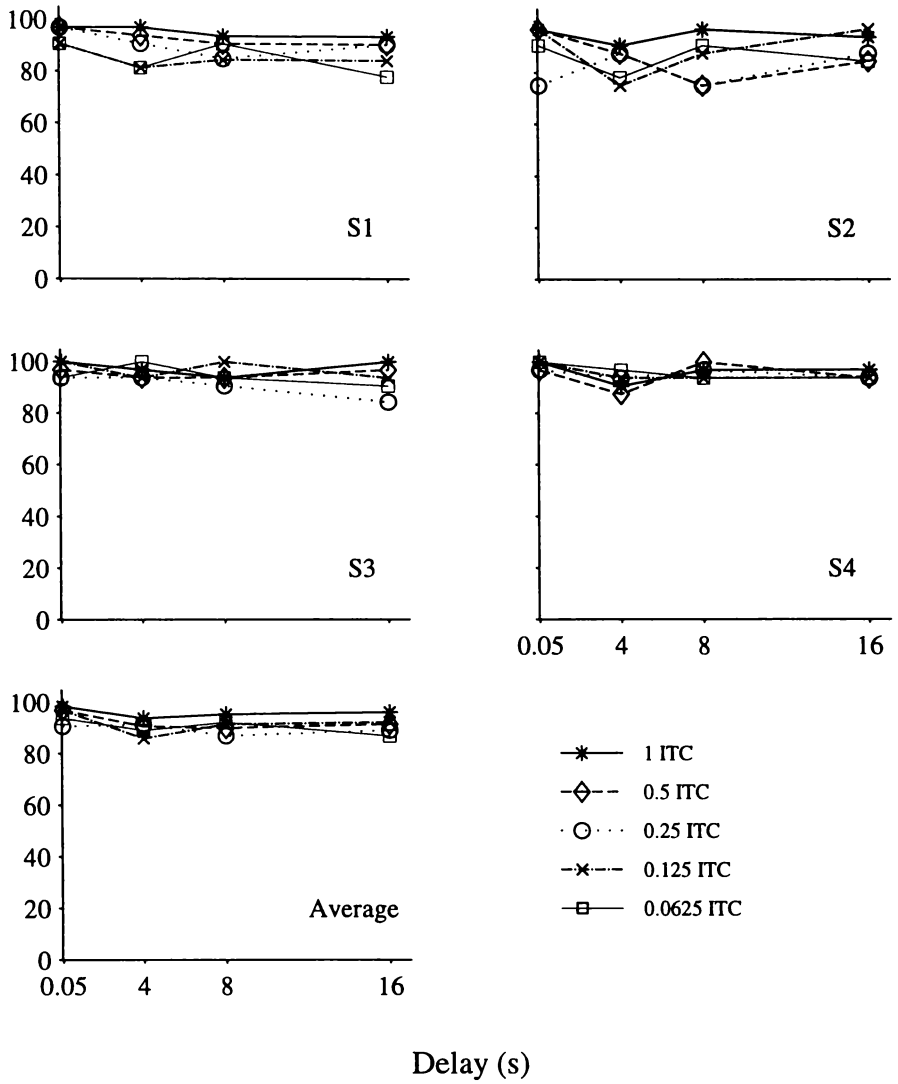

Figure 7.1. Percent correct for each participant and the average of these, plotted as a function of delay and the probability of intertrial correspondence.

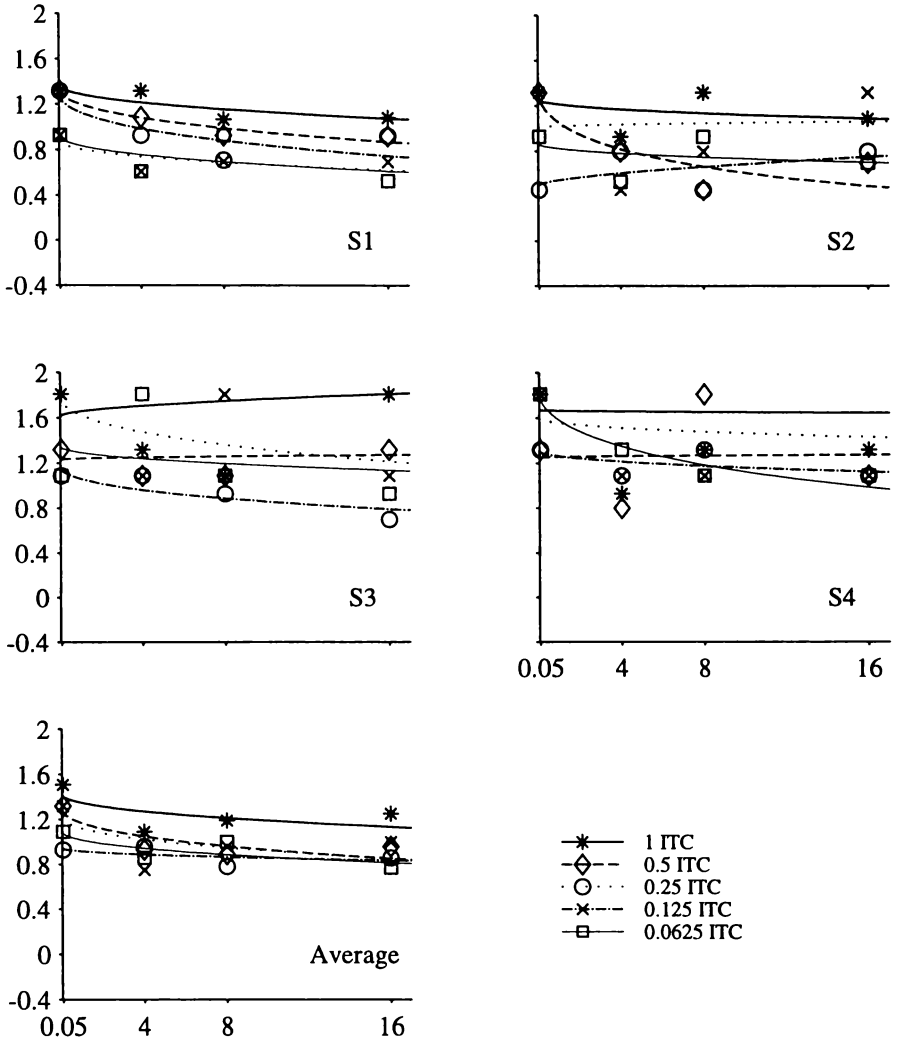

Figure 7.2. Estimates of discriminability (logit p) for the individual participants and the average of these, plotted as a function of delay. The data are fitted by negative exponential functions (Equation 2).

there was no systematic change in accuracy as the probability of ITC was decreased.

A repeated measures ANOVA was carried out on the logit *p* data. Prior to the ANOVA Mauchly's test of sphericity was used to test for homogeneity of variance. It was found that this was not significant for delay (W = 0.07) but was significant for the probability of ITC (W = 0.00) and the interaction (W = 0.00) ($\alpha<0.05$). Given this the Huynh-Feldt (1970) correction was used when testing for significance of the ANOVA. For the ANOVA the main effect of delay ($F = 9.38$, df = 3.00, 9.00) was statistically significant ($\alpha < 0.05$) but the main effect of ITC ($F = 3.62$, df = 3.03, 9.08) and the interaction ($F = 0.74$, df = 12.00, 36.00) were not statistically significant.

Negative exponential functions (Equation 2) were fitted to the data for each of the five conditions using non-linear regression and are shown on Figure 7.2. Table 7.1 shows the values obtained for the parameters *a* and *b* and the standard error of estimate and the percentages of VAC for by the fitted functions for both the individual and the averaged data. Accuracy generally decreased slightly as the delay was increased as shown by the *b* values in Table 7.1. There was no systematic effect on *b*, for any of the participants or for the averaged data, as the probability of ITC was decreased. For S1 there was a systematic tendency for *a* to decrease, with the exception of Condition 2, (where the probability of ITC was .0625) as the probability of ITC was decreased. For the remaining three participants there was no systematic change in *a* as the probability of ITC was decreased. For the averaged data there was a systematic tendency for *a* to decrease, with the exception of Condition 5, (where the probability of ITC was 0.25) as the probability of ITC was decreased. A unidirectional Ferguson test for trend (Ferguson, 1965) showed that *a* did not trend as the probability of ITC was decreased ($\alpha < 0.05$). A unidirectional Ferguson test for trend (Ferguson, 1965) also showed that *b* did not trend as the probability of ITC was decreased ($\alpha < 0.05$).

The data were analysed to compare performance on corresponding trial pairs and on non-corresponding trial pairs. Logit *p* was calculated for both the corresponding and the non-corresponding trials for each of the four participants as

Table 7.1 *Estimates of a, b, (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions for each of the experimental conditions.*

	ITC	*a*	*b*	VAC%	Std Err
S1	1	1.37	0.06	67	0.07
	0.5	1.34	0.11	93	0.04
	0.25	1.30	0.14	65	0.13
	0.125	0.88	0.09	51	0.08
	0.0625	0.93	0.10	36	0.14
S2	1	1.26	0.04	11	0.16
	0.5	1.36	0.25	78	0.15
	0.25	0.49	-0.11	28	0.14
	0.125	1.01	0.01	0	0.37
	0.0625	0.87	0.05	11	0.16
S3	1	1.61	0.03	3	0.31
	0.5	1.23	0.01	1	0.12
	0.25	1.17	0.10	71	0.08
	0.125	1.79	0.10	28	0.31
	0.0625	1.35	0.04	5	0.33
S4	1	1.67	0.10	31	0.26
	0.5	1.25	0.00	0	0.37
	0.25	1.31	0.04	29	0.10
	0.125	1.79	0.17	81	0.14
	0.0625	1.85	0.16	94	0.07

Average	1	1.43	0.06	39	0.12
	0.5	1.29	0.11	70	0.09
	0.25	0.94	0.03	29	0.06
	0.125	1.23	0.09	34	0.17
	0.0625	1.09	0.07	64	0.07

a function of delay and these are shown in Figure 7.3. For the individual and the averaged data logit *p* tended to decrease as the delay interval was increased and was, with the exception of S4, higher for corresponding than for non-corresponding trial pairs.

Negative exponential functions (Equation 2) were fitted to the data using non-linear regression. Table 7.2 shows the values obtained for the parameters *a* and *b* for both the individual and the averaged data. Table 7.2 also shows the standard error of estimate and the percentage of VAC by the fitted functions. For all of the participants and for the averaged data *a* was higher for corresponding trial pairs than it was for non-corresponding trial pairs. There were no systematic differences in *b* as a function of whether trial pairs were corresponding or non-corresponding.

The data were analysed to examine whether the non-matching comparison stimulus on trial *n* had appeared as the sample stimulus on trial *n-1*, for any of the individual participants. In the present experiment the size of the non-matching comparison stimuli for S1 and S2 were 55 and 65 pixels for the smaller of the two sample stimuli and 57 and 69 pixels for the larger of the two sample stimuli. For S3 and S4 the non-matching comparison stimuli were 53 and 67 pixels for the smaller of the two sample stimuli and 56 and 70 pixels for the larger of the two sample stimuli. This meant that the non-matching comparison stimulus on trial *n* had never appeared as the sample stimulus on trial *n* - 1 for any of the four participants.

An analysis of response bias for a particular alternative, based on the size of the comparison stimuli, was carried out for the individual participants as a function of delay and the probability of ITC. Response bias was calculated as the logarithm of the ratio of the total number of times that the small alternative was selected, during the choice phase, to the total number of times the large alternative was selected. There was no systematic tendency to choose either the larger or the smaller of the two comparison stimuli and there was no systematic change in this bias as the probability of ITC was decreased. As there was no bias, based on the size of the comparison stimuli these data are not shown here.

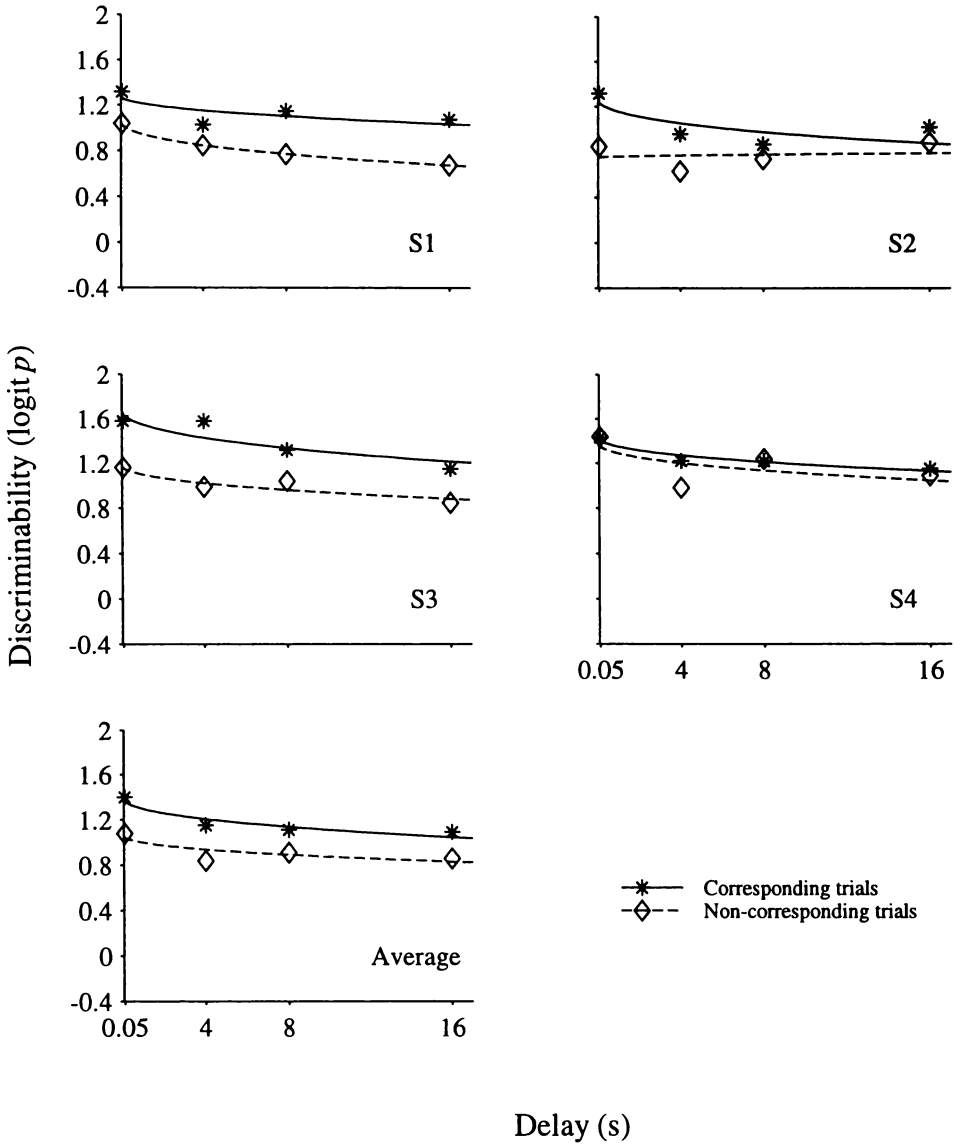

Figure 7.3. Estimates of discriminability (logit p) for the individuals and the average of these for corresponding and non-corresponding trials as a function of delay. The fitted functions are negative exponentials (Equation 2).

Table 7.2 *Estimates of a, b, (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions as a function of ITC across all experimental conditions.*

	ITC	*a*	*b*	VAC%	Std Err
S1	CT	1.28	0.05	54	0.07
	NCT	1.06	0.11	100	0.00
S2	CT	1.27	0.09	59	0.11
	NCT	0.75	-0.01	2	0.10
S3	CT	1.68	0.08	76	0.09
	NCT	1.18	0.07	84	0.04
S4	CT	1.43	0.06	93	0.03
	NCT	1.38	0.07	43	0.13
Average	CT	1.39	0.07	88	0.04
	NCT	1.06	0.06	67	0.05

An analysis of response bias for a particular alternative, based on the side of the screen that the comparison stimuli appeared on, was also carried out for all of the participants as a function of the probability of ITC and delay interval. These data are not shown here, as there was no systematic bias towards selecting either the comparison stimulus on the left or on the right of the screen and this bias did not change systematically as a function of the probability of ITC.

Figure 7.4 shows the percent correct for Condition 1 and Condition 6, where the probability of ITC was 1.0, for each of the four participants and the average of these, as a function of delay. With the exception of S2 there was no systematic difference in accuracy between Conditions 1 and 6 for the individual or the averaged data. For S2 accuracy was slightly higher for Condition 6 than for Condition 1.

The individual data were also analysed for any change in performance as the experimental session progressed. Each of the experimental conditions was divided into two blocks of 16 trials for each of the four delay intervals. Matching accuracy for the first block of 16 trials was then compared to matching accuracy for the second block of 16 trials, as a function of delay interval. There was no systematic change in matching accuracy across the first and second blocks for any of the experimental conditions.

Discussion

In the present study, accuracy decreased as the delay interval was increased, but the decease was small as shown by the generally small *b* values for the fitted exponential functions. There was no systematic change in these *b* values as the probability of ITC was decreased. There was also no systematic change in initial discriminability as the probability of ITC was decreased, as shown by the *a* values for the fitted exponential functions. Thus, for the present experiment, decreasing the probability of ITC did not change either *a* or *b*. This suggests that the change in discriminability which occurred as the number of sample stimuli was increased, previously, was not a result of the corresponding increase in the number of non-corresponding pairs of consecutive trials.

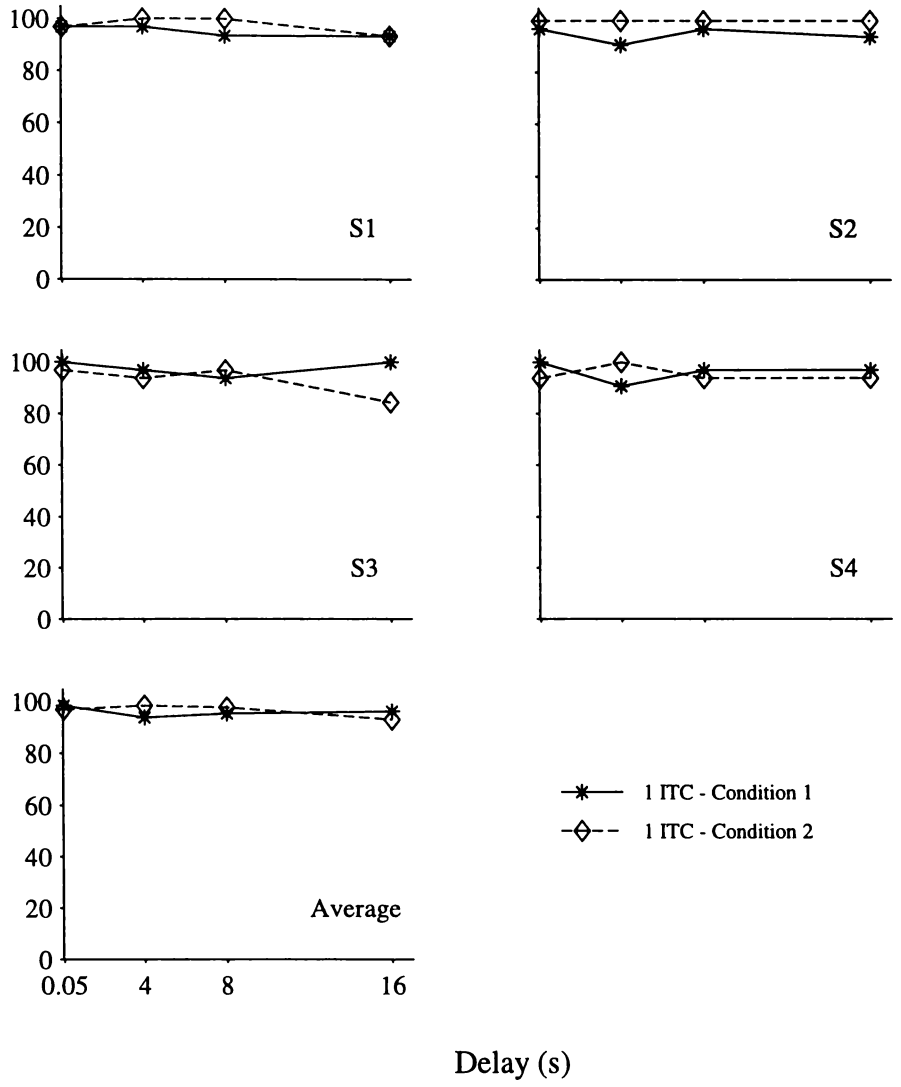

Figure 7.4. Percent correct for each participant and the average of these, for Conditions 1 and 6, as a function of delay.

Despite the fact that accuracy did not decrease systematically as the probability of ITC was decreased there was still a difference in accuracy on corresponding and non-corresponding trials over all of the experimental conditions (Figure 7.3). Generally, *a* was higher on corresponding trials than on non-corresponding trials.

In the previous experiments performance on corresponding and non-corresponding trials was a product of performance on trials with more than one of the sample-set sizes. In the case of the present experiment, corresponding and non-corresponding trials were a product of performance with either one or two sample stimuli. In addition, a large proportion of the corresponding trials were from the one-sample condition and a large proportion of the non-corresponding trials were from the two-sample condition. Thus, the overall difference, here, between accuracy on corresponding and non-corresponding trials may reflect the effect that changing the number of sample stimuli was found to have on accuracy in previous experiments, independent of the effect of ITC.

To see how much the one-sample condition contributed to the different accuracy on corresponding and non-corresponding trials the individual data, here, were re-analysed excluding trials from the one-sample condition. Figure 7.5 shows logit *p* for each of the four participants and the average of these for corresponding and non-corresponding trials, with the exclusion of trials from the one-sample condition. There was generally little or no change in the level of accuracy obtained for the corresponding trails when the one-sample trials were excluded.

Negative exponential functions (Equation 2) were fitted to the data using non-linear regression and are shown on Table 7.3. There was no change in the tendency for *a* to be higher for corresponding than for non-corresponding trials. Thus, the one-sample condition does not account for the present finding that accuracy was higher on corresponding than on non-corresponding trials.

In Experiments 5 and 6 a one sample titration was used. This meant that, there, accuracy on the one-sample condition tended to be high and tended to decreased as the number of sample stimuli was increased. Since the present experiment used a two-sample titration it might be expected that accuracy

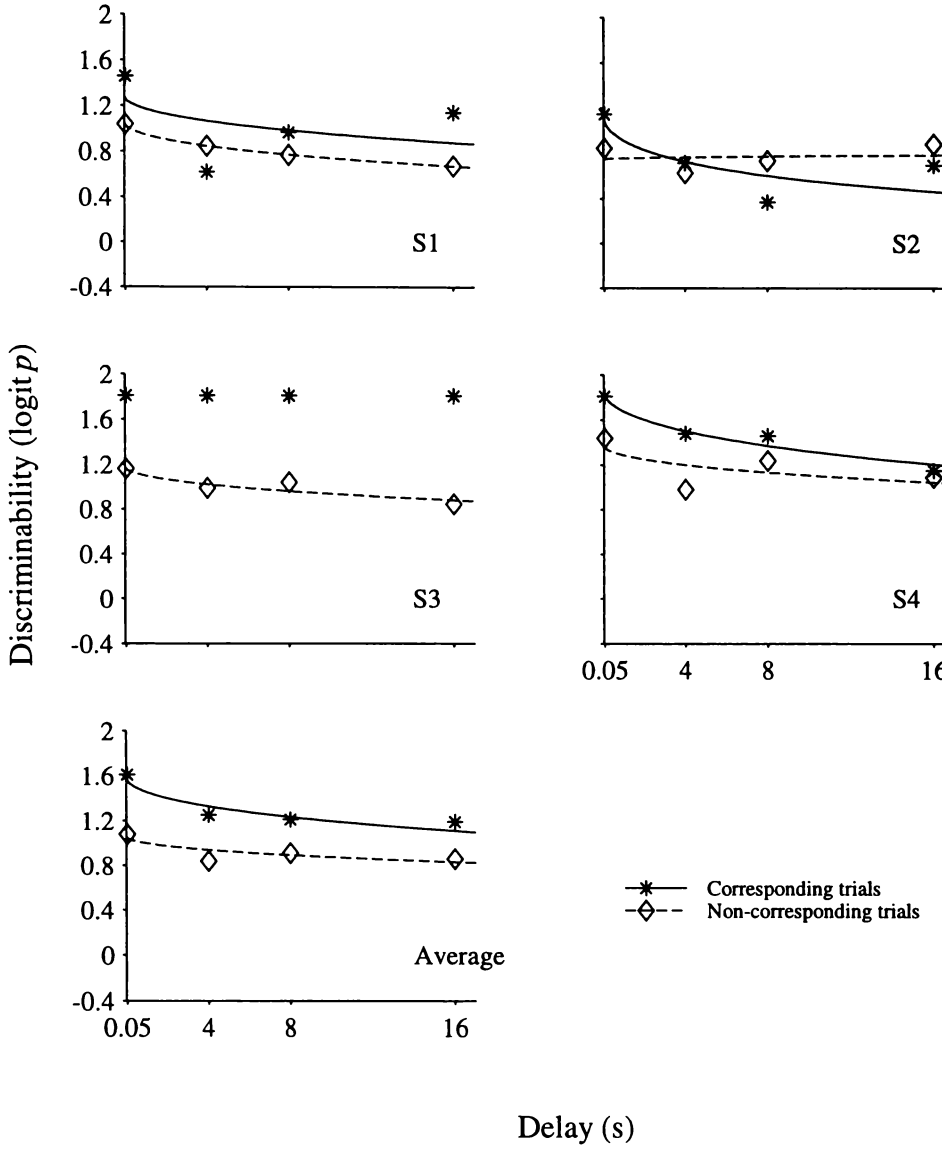

Figure 7.5. Estimates of discriminability (logit p) for the individuals and the average of these for corresponding and non-corresponding trials, excluding the One-sample Condition, as a function of delay. The fitted functions are negative exponentials (Equation 2).

Table 7.3 *Estimates of a, b, (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions as a function of ITC across all experimental conditions, excluding the one-sample condition.*

	ITC	*a*	*b*	VAC%	Std Err
S1	CT	1.29	0.09	17	0.28
	NCT	1.14	0.12	100	4.79
S2	CT	0.79	-0.01	2	0.12
	NCT	1.15	0.23	63	0.17
S3	CT	*	*	*	*
	NCT	1.33	0.08	85	0.05
S4	CT	1.87	0.11	94	0.06
	NCT	1.61	0.09	48	0.17
Average	CT	1.60	0.09	87	0.19
	NCT	1.14	0.07	68	0.06

* Indicates that the negative exponential function could not be fitted to the data for this participant for corresponding trials.

(Condition 1) with only one sample stimulus would be higher than with two sample stimuli. Thus, it is possible that for the present two sample procedure titrating to a criterion of 100% led to a ceiling effect, which may have masked any possible effect of the decrease in ITC. This suggests that to further investigate the effect of ITC on matching accuracy it would be necessary to use a criterion level of accuracy, for the titration, which was lower than 100% correct.

As mentioned above the previous multiple-sample experiments titrated to 100% using a one-sample titration. This meant that any cases in which a participant was able to achieve very high levels of accuracy, that is close to or 100% correct, during the DMTS task were generally for the one-sample condition. However, in the present experiment accuracy was generally very high for all of the participants, for all of the experimental conditions.

The aim of the present study was to investigate the impact of decreasing the probability of ITC on matching accuracy. The present study only used two sample stimuli, during the DMTS task, as did Edhouse and White (1988a) when investigating the overall effect of corresponding and non-corresponding trials on matching accuracy with pigeons. However, unlike Edhouse and White's (1988a) procedure the sample stimuli did not appear as the non-matching comparison stimuli and thus the present task was not a conditional discrimination. Given that Edhouse and White's (1988a) is the only study to investigate the impact of corresponding and non-corresponding trials on matching accuracy directly, the next step might be to use a procedure similar to the one which they used. Thus, in order to investigate the impact of ITC on matching accuracy further it might be better to use a traditional two-sample procedure as used by Edhouse and White (1988a).

In conclusion the present result suggests that the decrease in accuracy which occurred as the number of sample stimuli was increased, in the previous experiments, was not due to the corresponding decrease in the probability of ITC. However, given the possible ceiling effect it is not clear whether this result was a true indication of how accuracy would be effected as the probability of ITC was decreased.

Experiment 8

It was suggested in the previous experiment that a traditional two-sample conditional discrimination task, similar to the one used by Edhouse and White (1988a), should be used to investigate further whether decreasing the probability of ITC results in a decrease in *a*. This type of two-sample DMTS task has not traditionally been used with human participants as it has been suggested that the stimuli could be labelled, and hence rehearsed during the delay (Parr, 1992). It is not clear whether such a typical two-sample conditional discrimination task, using two different circles as both sample and comparison stimuli can be used with human participants. To see if this was possible, the present study used such a task. Whereas previously the data were analysed using logit *p* the move to a traditional two-sample task also allows the calculation of log *d*, a bias free measure of discriminability, and log *b*, a measure of bias independent of discriminability.

This next experiment also aimed to investigated whether decreasing the physical disparity between the two circular stimuli would make discrimination more difficult and so result in a decrease in initial discriminability, *a*, as found by White (1985). If initial discriminability decreased as the physical disparity decreased it would be possible to change the difficulty of the task and use a titration procedure similar to that used previously to obtain similar performance across participants. Thus, the difficultly of the task could be manipulated by changing the physical disparity between the two sample stimuli. This would be an advantage as in the previous experiment accuracy was very high in all conditions and it was suggested that a ceiling effect may have masked any effect of ITC on discriminability. To overcome this issue, when further studying the effect of ITC on matching accuracy, it was suggested that a titration with a criterion level of accuracy of less than 100% correct should be used.

In the present study a traditional two-sample conditional discrimination was used, with a series of four conditions and for each of the four conditions the physical disparity between the two-sample stimuli was different. Two delays were used with each of the different physical disparities to establish whether accuracy would decrease as the delay was increased, for this task.

Method

Subjects

Six first year psychology students participated in this experiment. Each participant received 1% course credit towards a first year psychology course, irrespective of their performance on the experimental task and irrespective of whether they completed the experiment.

Apparatus

The apparatus used here was the same as that used for Experiment 1.

Stimuli

The stimuli were filled white disks, presented on a blue background, ranging in size from 72 to 88 pixels in diameter, in steps of two pixels (making a total of eight different sized disks). In each condition two of the disks were used as sample stimuli. The pairs of sample stimuli were 78 and 82, 72 and 88, 76 and 84, and 74 and 86 pixels in diameter.

Procedure

The general procedure used in the present experiment was similar to that used in the previous experiments. However, the procedure used differed from those used previously in that there was no initial titration. The DMTS task used in here also differed in that only two delay intervals (0.05 s, and 4 s) were used. There were two sample stimuli in each condition and both were used as the comparison stimuli during the choice phase. The comparison disk which was the same size as the sample stimulus was the matching comparison stimulus and the comparison disk which was not the same size as the sample stimulus was the non-matching comparison stimulus.

For each trial, the delay interval, the order in which the sample stimuli appeared and the side on which the comparison stimuli were presented was pre-arranged using random selection. The same order was then used for all participants, thus the experimental events occurred in the same order for all of the participants.

There were five experimental conditions in which the physical disparity, measured in pixels in diameter, between the two sample disks was varied. In Conditions 1 to 4 the physical disparity between the diameters of the sample stimuli was 4, 16, 8, and 12 pixels, respectively. Condition 5 was a replication of Condition 1 and thus, there was a physical disparity of 4 pixels between the sample stimuli. Each sample stimulus was presented 30 times with each of the delay intervals and a total of 60 times for each condition. Each condition ended after the completion of 120 trials and the next condition began immediately. The experimental session ended when the participant had completed 600 trials at the conclusion of Condition 5. Once the experimental session had ended the participants were given details about the purpose of the experiment.

Results

Figure 8.1 shows percent correct for the six individual participants and for the averaged data, for each of the physical disparities, for both delays. With the exception of S3 and S5 accuracy was lowest for both delays when the sample stimuli differed in diameter by 4 pixels. Accuracy was generally much higher with the 8 pixel difference than with the 4 pixel difference with the exception of S3 and the 4 s delay for S5. There was generally little or no change in accuracy for any individual or for the average as the difference between the two sample stimuli was increased from 8 to 12 and to 16 pixels in diameter (Conditions 3, 4 and 2, respectively). Accuracy for S3 showed little or no change as the disparity between the sample stimuli was increased over the range of disparities used here.

Figure 8.2 shows log *d*, calculated using Equation 6 (Appendix A), for the individual participants, and for the averaged data, for each of the differences in diameter at each delay. These data show the same patterns as the percent correct data although the differences in accuracy between Conditions 2, 3, and 4 are clearer due to the ratio scale. Log *d* was generally higher for the 0.05 s delay than for the 4 s delay. Figure 8.2 also shows the data from the replication (Condition 5) of the four pixel condition (Condition 1). There was generally little or no change in matching accuracy over the two conditions and the differences that did occur were not consistent across all of the participants.

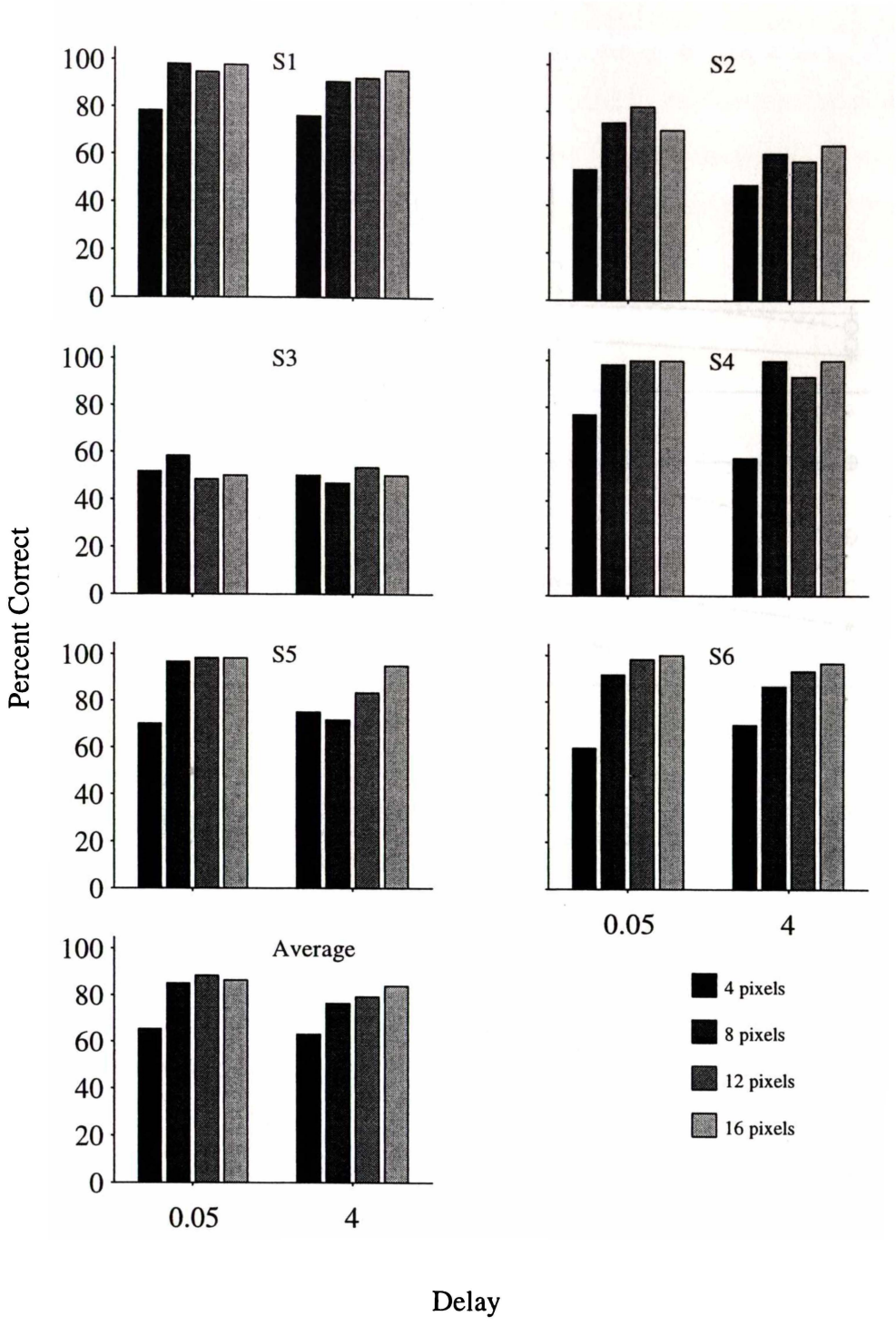

Figure 8.1. Percent correct for each of the diameter differences for each of the delay intervals for the individuals and the averaged data.

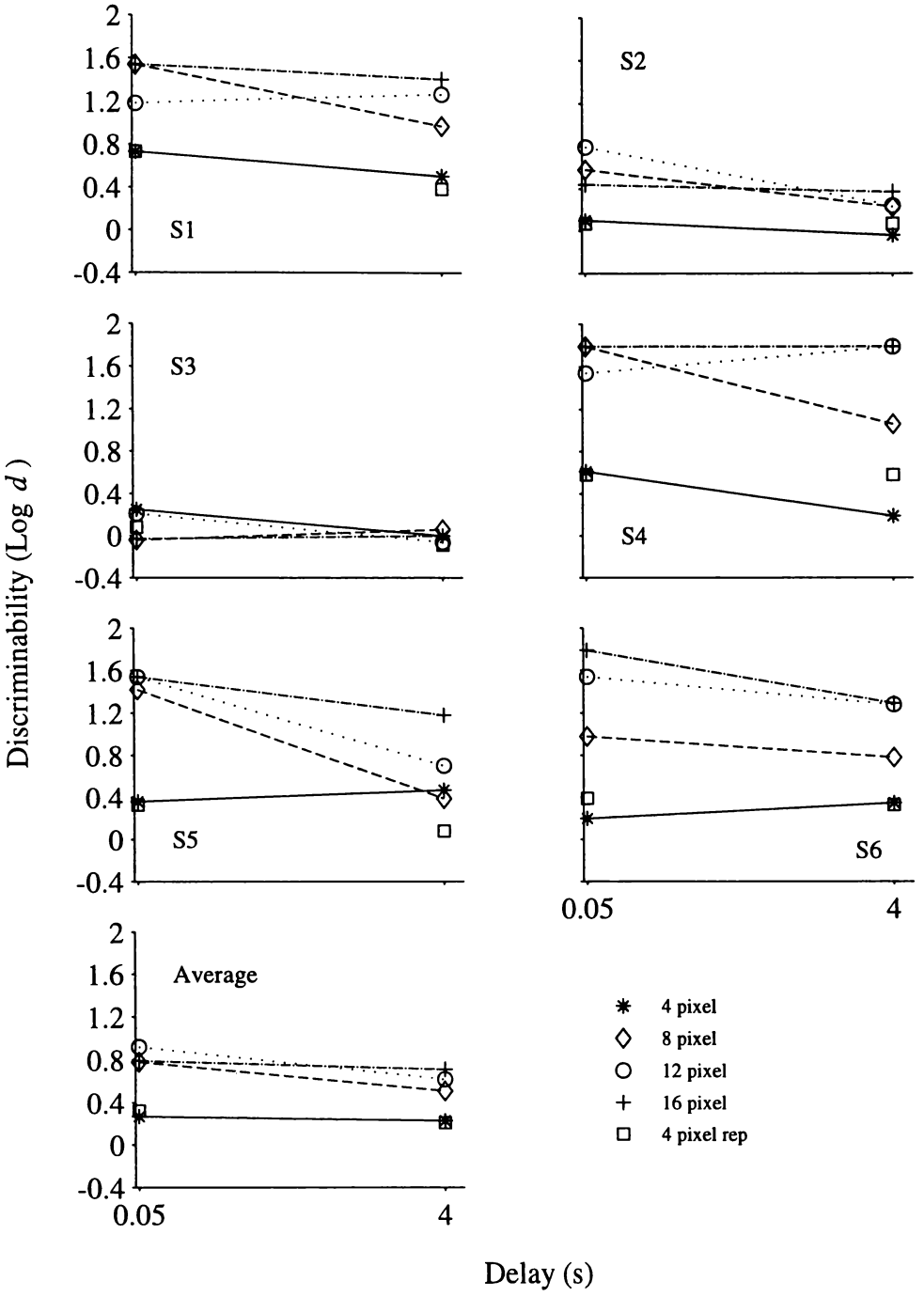

Figure 8.2. Log *d* plotted for each physical disparity and each delay interval for the individual participants and the averaged data.

A repeated measures ANOVA was carried out on the log *d* data. Prior to the ANOVA a Mauchly's test of sphericity was used to test for homogeneity of variance. It was significant for physical disparity (W = 0.06), delay (W = 1.00) and for the interaction (W = 0.65) ($\alpha = 0.05$). Thus the Huynh-Feldt (1970) correction was used when carrying out the ANOVA. For the ANOVA the main effect of physical disparity (F = 10.61, df = 1.43, 7.13) and delay (F = 18.19, df = 1.00, 3.00) were statistically significant and the interaction (F = 1.20, df = 3.00, 15.00) was not statistically significant ($\alpha > 0.05$). A pairwise comparison of means was carried out using Tukey's LSD. For the main effect of physical disparity, the mean of Condition 1 was significantly different from the means of Conditions 2, 3 and 4 but the means of Conditions 2, 3 and 4 were not significantly different from each other ($\alpha > 0.05$).

An analysis of response bias, log *b* (Equation 8, Appendix A), for a particular alternative based on the size of the comparison stimuli as a function of the difference in the diameters of the two sample stimuli and of the delay was carried out for each of the individuals and for the averaged data. Figure 8.3 shows that for three out of the six participants (S2, S3 and S5) there was a bias towards selecting one of the two alternatives for at least three of the conditions for both delays. For these three participants this bias was generally not large, with the exception of Condition 1 for S3, and did not change systematically as a function of the disparity between the two sample stimuli. The remaining two participants showed little or no bias towards either of the alternatives. The average data were representative of the individual data in that there was a small bias towards one of the two alternatives but this bias did not change systematically as a function of physical disparity.

The data were also examined for any tendency to respond more to the comparison stimulus on the left or the right side of the screen. These data are not shown as there was little or no bias towards either of the two alternatives for any of the individuals or for the averaged data and that there was no change in bias as a function of physical disparity.

The individual data were examined for any systematic change in matching accuracy across the experimental session. Each of the experimental conditions

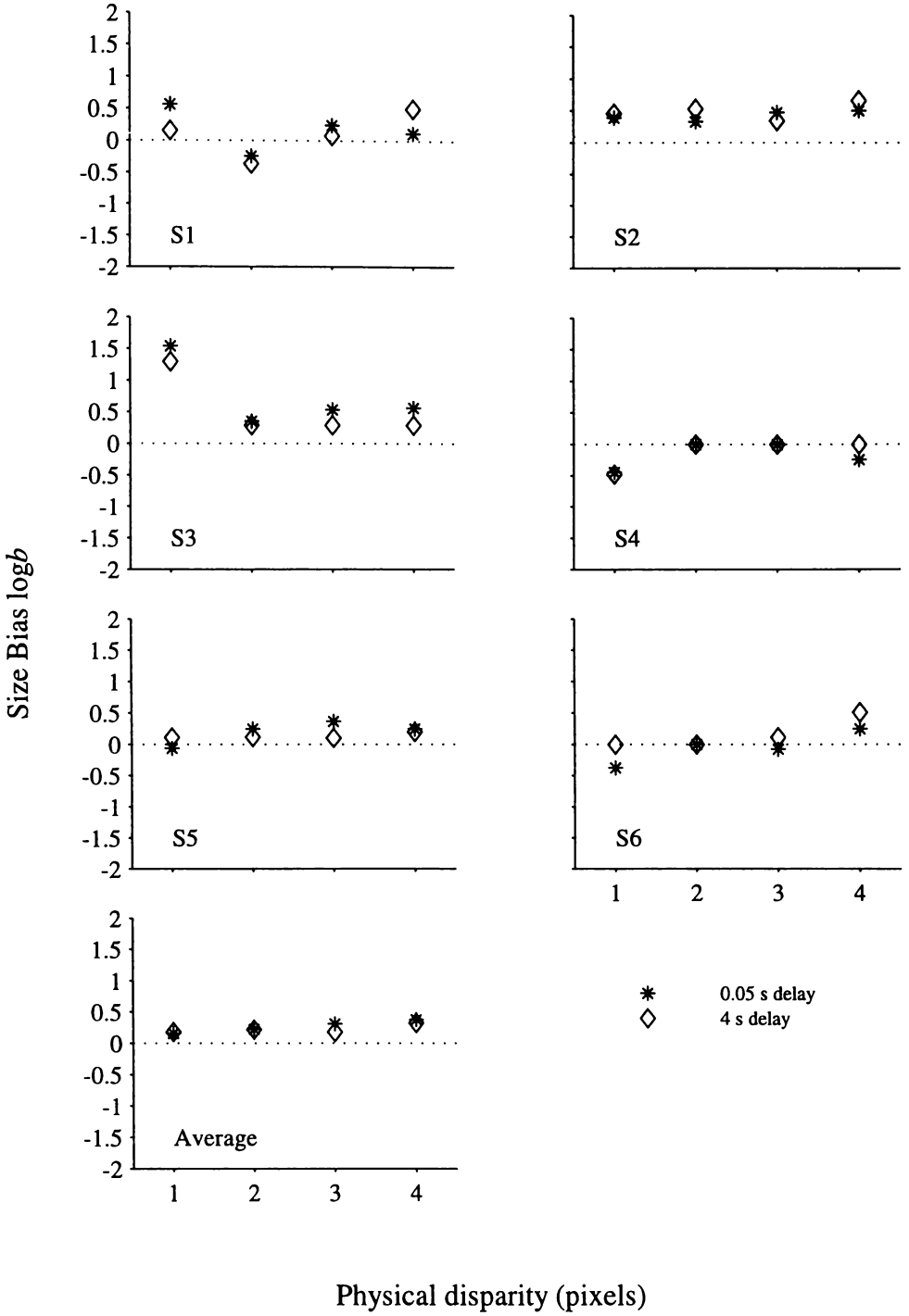

Figure 8.3. Log *b*, response bias plotted as a function of the size of the comparison stimuli, for each physical disparity for the individual and the averaged data.

was split into two blocks of 60 trials for each of the delay intervals. Matching accuracy for the first block of trials was compared to matching accuracy for the second block of trials as a function of delay. These are not presented as there was no systematic change in matching accuracy across the two blocks of experimental trials. This was consistent with the finding that there was no systematic change in accuracy over Condition 1 and Condition 5.

Discussion

It was shown here that increasing the physical disparity between the sample stimuli generally increased matching accuracy, a finding which is consistent with that of White (1985). The present result suggests that the difficulty of the present two-sample DMTS task could be changed by changing this physical disparity. The results here also showed that accuracy for participants was generally only high when the physical disparity between the two stimuli was large. As noted earlier it has been suggested that when a two sample DMTS task was used with human participants that it would be possible for the participants to verbally encode (Parr 1992) the sample stimuli and that accuracy would be high. Thus, the present experiment suggests that if high accuracy with a two sample DMTS task with human participants is a result of verbal encoding that for the present task verbal encoding was only possible when the physical disparity was large.

Individual participants performed at different levels of accuracy at the same physical disparity and thus, the task appeared to be harder for some participants than for others. This can be clearly seen for S3, whose accuracy was generally much lower than that of the other participants and for S5, whose accuracy was very high and similar for three of the conditions with the 0.05 s delay. This suggests that titrating the disparity between the stimuli could be used to establish equivalent levels of performance for all participants and avoid floor and ceiling effects, as in previous experiments. For the previous experiments the titration adjusted the physical disparity between the sample and the non-matching comparison stimuli. However, for the present task a titration would have to adjust

the disparity between the two sample stimuli and so the titration used previously would have to be modified.

In the present experiment four different physical disparities were used and for most participants there was a large increase in accuracy when the difference between the diameter of the sample stimuli was increased from 4 to 8 pixels. There were however, generally only a small increase in accuracy as the physical disparity was increased from 8 to 12 to 16 pixels. This suggests that the increase of four pixels used here may have been too large to show gradual increases in accuracy as the physical disparity between the sample stimuli was increased. Thus, a titration for the present task should probably increase the physical disparity by a smaller amount than 4 pixels.

In conclusion, the present experiment suggests that it is possible to use a traditional two-sample conditional discrimination using disks of different sizes, with human participants. Accuracy for the same level of physical disparity differed across participants. This finding is consistent with Adamson (1995) and suggests that, as previously, a titration task should be used to give similar level of performance across participants.

Experiment 9

The results of the previous experiment showed that the difficulty of the present two-sample conditional discrimination could be changed by changing the difference between the diameters of the two sample stimuli. However, at the same level of physical disparity matching accuracy varied across participants. Thus, to equalise performance across participants it would be necessary to use a titration similar to those used previously. This next study investigated whether such a titration would work for the present two-sample DMTS task. It was suggested previously, that the criterion level of accuracy should be lowered from 100% correct to avoid the occurrence of a ceiling effect on accuracy. Thus, the criterion level used here was 90% correct. The present experiment investigated matching accuracy on the same two-sample DMTS task used in the previous experiment across four delays (0.05, 4, 8 and 16 s).

Method

Subjects

Five-first year psychology students participated in this experiment. Each participant received a one percent course credit for each hour of participation, irrespective of their performance on the experimental task.

Apparatus

The apparatus used here was the same as that used in Experiment 1.

Stimuli

The stimuli were two filled white disks, presented on a blue background. At the beginning of the experiment there was a difference of two pixels between the two stimuli and the stimuli were 98 and 100 pixels in diameter. The size of the sample disks were adjusted during the initial phase of the experiment for each of the participants.

Procedure

The general procedure was the same as that used in Experiment 2. The titration procedure used here differed in that both of the sample stimuli appeared as comparison stimuli on every experimental trial. The titration procedure also differed in that the physical disparity between the two initial sample disks was increased until the participant reached the criterion level of accuracy, 90% correct. The two sample disks initially differed in diameter by two pixels. At the completion of the titration the DMTS phase of the session began, using the size of the sample stimuli reached at the completion of the titration. All of the experimental events in the titration procedure were randomly selected and controlled by the computer.

The general procedure used for the DMTS phase of the experiment was the same as that used in Experiment 8 except for the inclusion of one of four delays (0.05, 4, 8, or 16 s) between the presentation of the sample stimulus and the choice phase. The delay intervals appeared in blocks of 32 trials, in the following pseudo-random order 4, 8, 16, and 0.05 s. The experimental events occurred in the same order for all of the participants. There was a total of 256 trials for the DMTS procedure with a total of 64 trials at each delay. The DMTS task was divided into two blocks of 128 trials, with 32 trials at each of the delay intervals. The DMTS session ended when the participant had completed 256 trials, in total, at the end of the second block of the DMTS session. Once the experimental session had ended the participants were informed about the purpose of the experiment.

Results

S1 to S5 completed the titration in 70, 110, 140, 160 and 130 trials with differences of 10, 14, 20, 24, and 18 pixels in diameter between the two stimuli, respectively. Thus S4 entered the main part of the session with the largest physical disparity and S1 entered with the smallest physical disparity, relative to the other participants.

Figure 9.1 shows percent correct for each of the five participants and for the average data as a function of delay interval. Accuracy was high for all five participants for the 0.05 s delay and with the exception of S3, accuracy was high with all other delays. Overall, there was a slight decrease in accuracy as the delay interval was increased, with the exception of S1. The average data are similar to the individual data in that accuracy was high for the delay of 0.05 s and decreased slightly as the delay interval was increased.

Figure 9.2 shows log *d*, calculated using Equation 6 (Appendix A), for each of the five participants and for the averaged data as function of delay interval. Negative exponential functions (Equation 2) were fitted to the log *d* data using non-linear regression. Table 9.1 shows the values obtained for the parameters *a* and *b* for both the individual and the average data. Table 9.1 also shows the standard errors of estimate and the percentages of variance accounted for (VAC) by the fitted functions. For all participants *a* was generally similar and high. With the exception of S1, there was a decrease in accuracy as the delay increased, for all participants, as shown by the *b* values in Table 9.1. For S1 accuracy tended to increase as the delay was increased. The standard error of estimate was generally small for all of the individuals and for the averaged data. The percentage VAC was generally large for the individual data and for the averaged data, with the exception of S1 and S5.

As for Experiments 5, 6 and 7 an overall analysis of corresponding and non-corresponding trials was carried out for the individual and the averaged data. Figure 9.3 shows log *d* for the individual and the average data for corresponding and non-corresponding trials for each delay. Negative exponential functions were fitted to the data and the parameters obtained for *a* and *b* are given in Table 9.2 for both the individual and the averaged data. Table 9.2 also shows the standard errors of estimate and the percentages of variance accounted for (VAC) by the fitted functions. Accuracy was generally high for both corresponding and non-corresponding trial types, as shown by the *a* values in Table 9.2. Accuracy tended to decrease as the delay interval was increased for both corresponding and non-corresponding trials, with the exception of S1 and non-corresponding trials

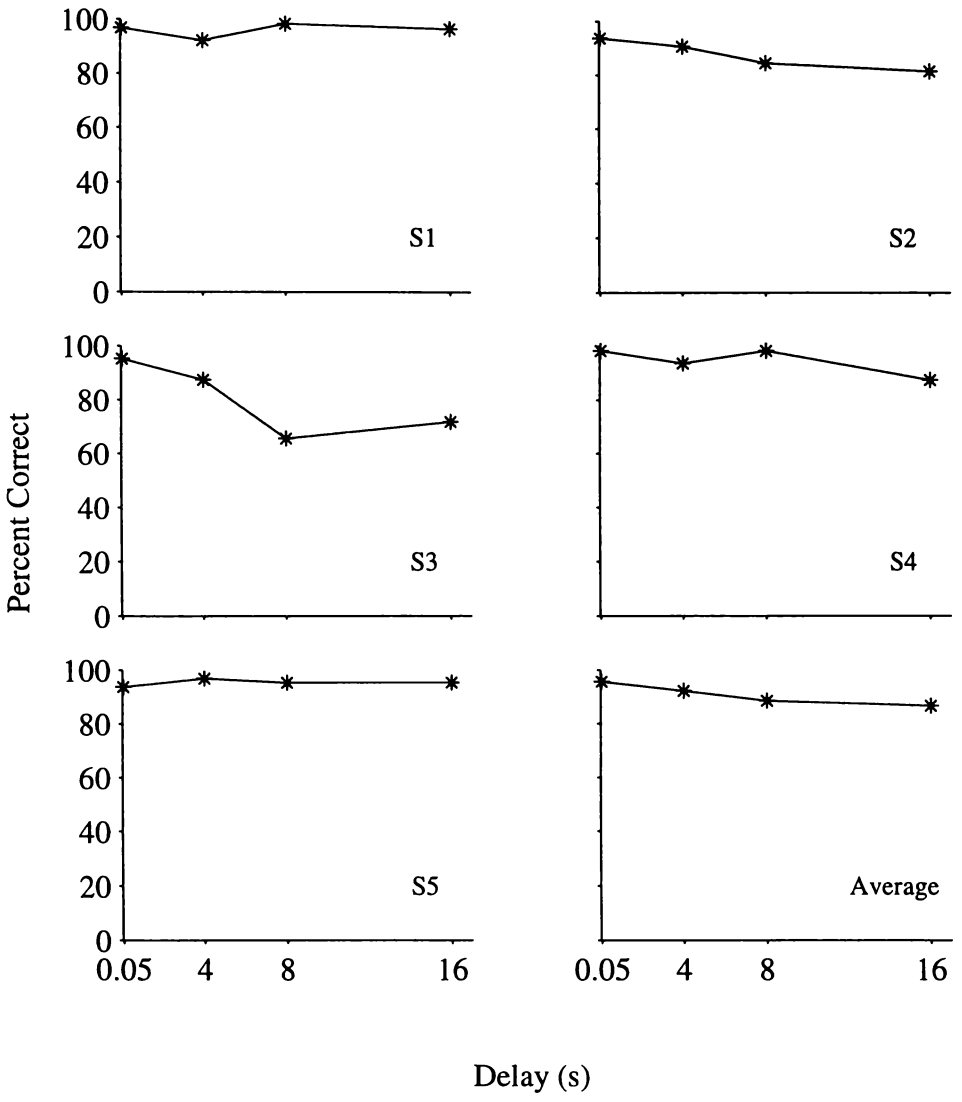

Figure 9.1. Percent correct for each participant and the average of these, plotted as a function of delay.

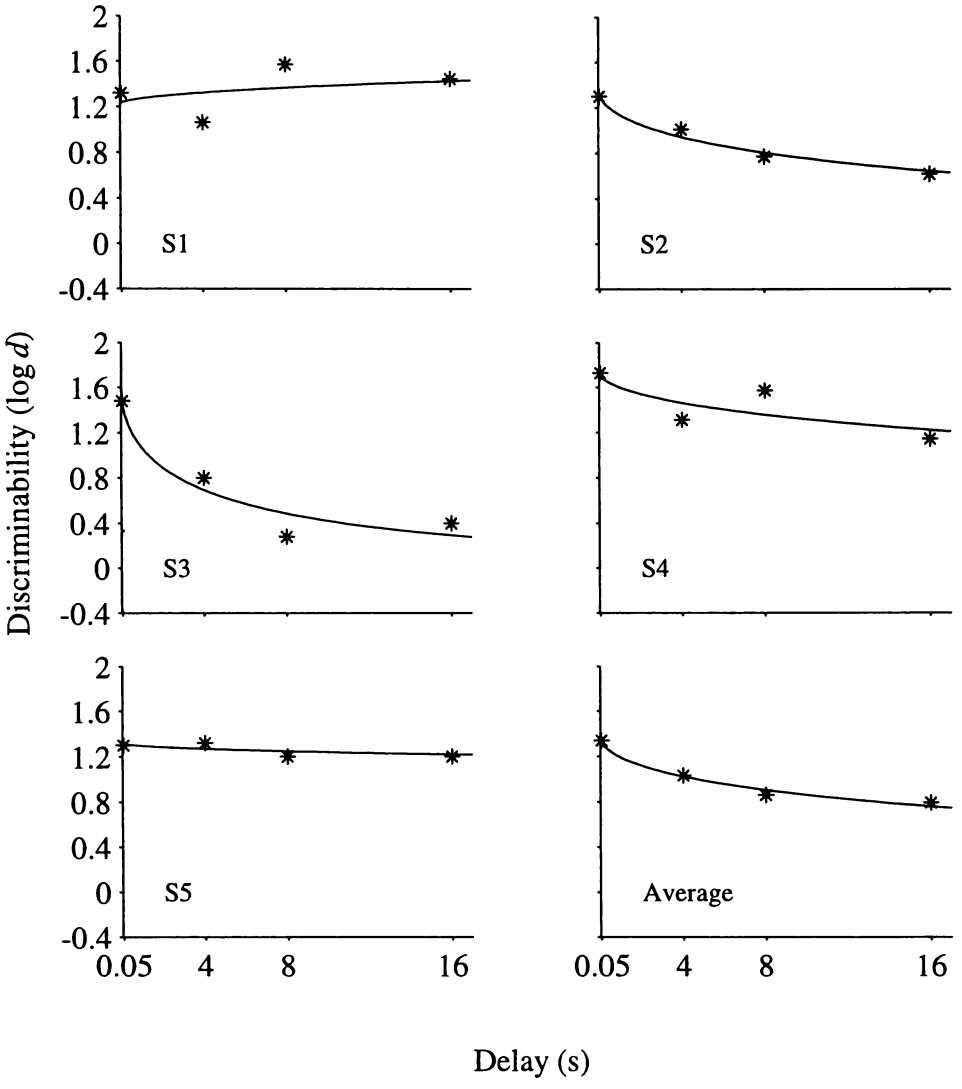

Figure 9.2. Discriminability (log *d*) for the individual participants and the average of these, plotted as a function of delay. The functions shown are negative exponentials (Equation 2).

Table 9.1 *a, b, (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions for the individuals and the average of these.*

	a	*b*	VAC%	Std Err
S1	1.22	-0.04	18	0.17
S2	1.38	0.19	97	0.04
S3	1.64	0.43	93	0.13
S4	1.75	0.09	64	0.14
S5	1.32	0.02	56	0.04
Average	1.38	0.15	98	0.03

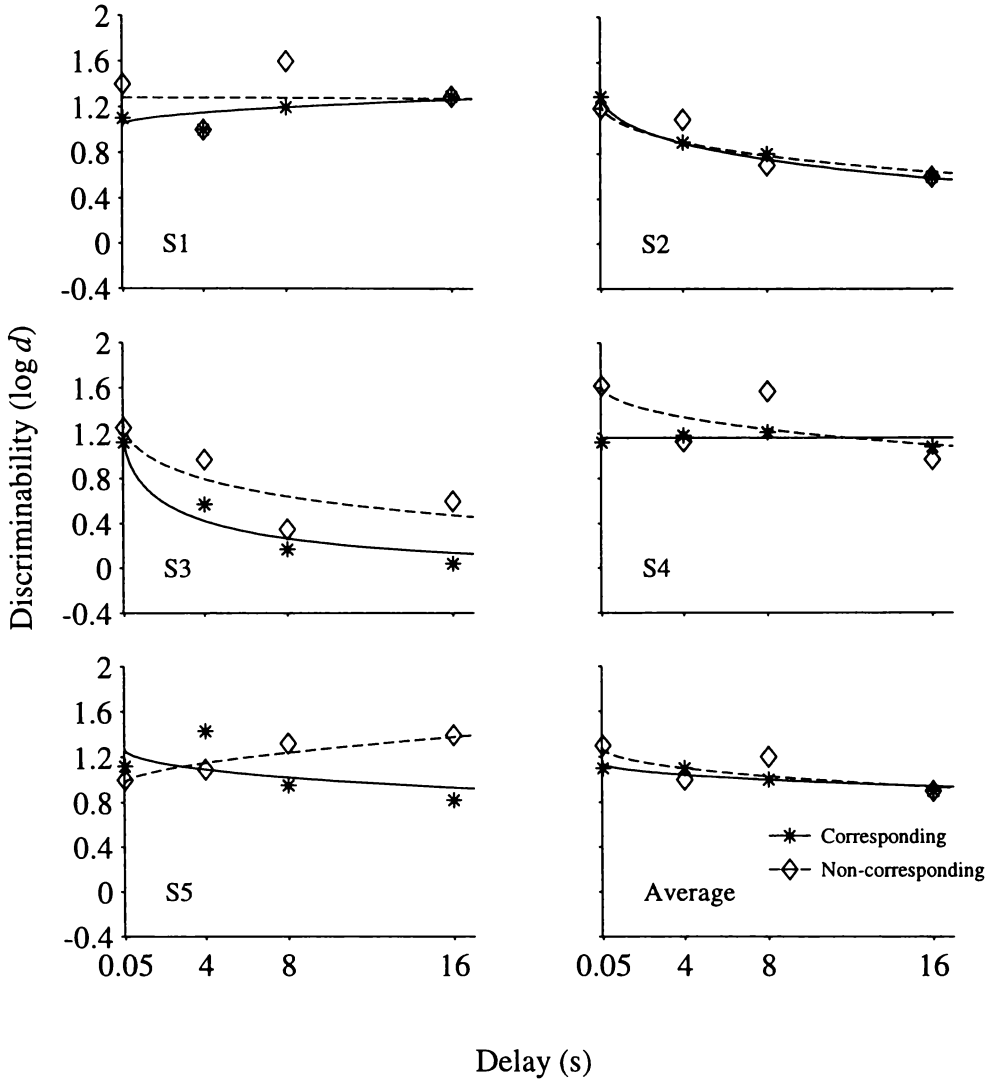

Figure 9.3. Discriminability (log *d*) plotted as a function of delay for corresponding and non-corresponding trials for each participant and the average of these. The fitted functions are negative exponentials (Equation 2).

Table 9.2 *a, b, (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions for corresponding and non-corresponding trials for each of the participants and the average of these.*

	Trial Type	*a*	*b*	VAC %	Std Err
S1	CT	1.04	-0.05	58	0.07
	NCT	1.28	0.00	0	0.22
S2	CT	1.36	0.21	98	0.04
	NCT	1.27	0.17	80	0.11
S3	CT	1.30	0.56	94	0.10
	NCT	1.34	0.26	72	0.18
S4	CT	1.16	0.00	1	0.05
	NCT	1.64	0.10	45	0.21
S5	CT	1.28	0.08	28	0.19
	NCT	0.96	-0.09	91	0.05
Average	CT	1.15	0.05	72	0.04
	NCT	1.29	0.08	73	0.07

for S5. There was no consistent difference between *a* or *b* as a function of whether trials were corresponding or non-corresponding across participants.

An analysis of response bias (log *b*) calculated using Equation 8 (Appendix A), based on the size of the comparison stimuli, was carried out for the individual participants and the averaged data as a function of delay. The response bias was calculated using Equation 8 (Appendix A). Figure 9.4 shows that there was generally a small bias towards selecting one of the two comparison stimuli more than the other but that this bias did not change in a systematic way as a function of delay. For four out of the five participants (S1, S2, S3 and S4) and the averaged data this bias was generally towards selecting the smaller of the two comparison stimuli (values > 0). For S5, there was generally a tendency towards selecting the larger of the two sample stimuli (values < 0).

Log *b* was also calculated, for both the individuals and the averaged data, as a function of the side of the screen on which the comparison stimuli had appeared and for each delay. The response bias was calculated using Equation 8 (Appendix A). These data are not shown as there was little or no bias towards selecting the comparison stimulus on either the left or the right of the screen for any of the individual participants or for the averaged data.

The data for each of the individual participants was analysed for any changes in performance across the experiment session. In order to do this the data, for each of the participants, from the first and the second phase of the experimental session was analysed separately as a function of delay interval. These data are not presented as there was no systematic change in accuracy across the two experimental blocks.

Discussion

The present study provides support for the previous suggestion that it is possible to use disks in a two sample conditional discrimination task similar to that used by Edhouse and White (1988a) with human participants. Initial accuracy was high for the shortest delay as shown by *a* for the fitted functions. With the exception of S1, there was typically some decrease in accuracy with

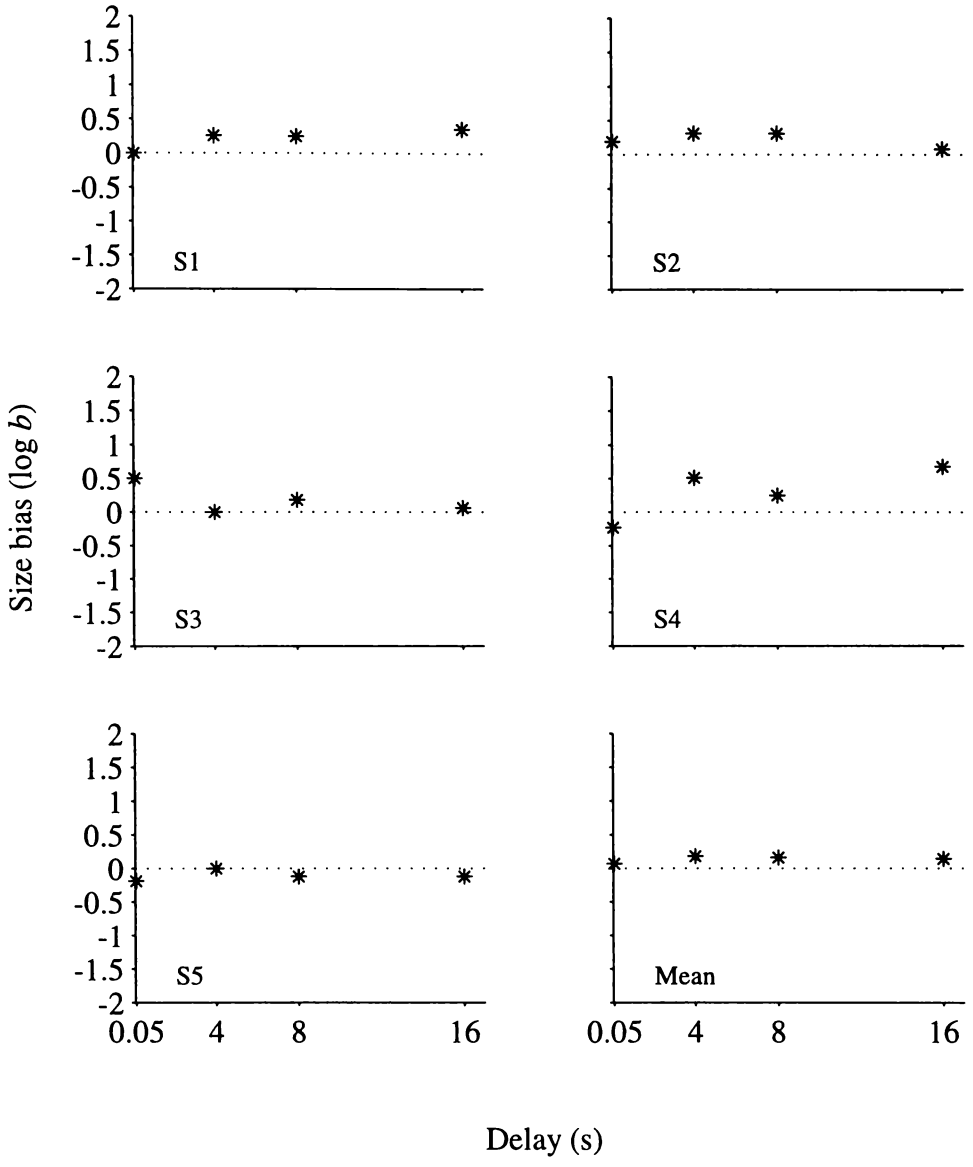

Figure 9.4. Response bias (log *b*) plotted as a function of the size of the comparison stimuli for each participant and the average of these.

increases in delay as shown by *b*. The present results showed that accuracy was generally consistent across participants for the delay of 0.05 s. This suggests that adjusting the physical disparity between the two sample stimuli for each of the participants resulted in similar levels of performance across the individuals. This finding, for this two-sample conditional discrimination, is consistent with those of previous studies here.

The difference in physical disparity between the sample stimuli reached at the end of the titration procedure ranged from 10 to 24 pixels in diameter across participants. In Experiment 8 most participants were performing at or above 90% correct with a difference of 8 pixels in diameter. Thus, the final physical disparities reached here (10 to 24 pixels), are generally larger than those in Experiment 8, for a similar level of accuracy.

In Experiment 7 where the affect of decreasing the probability of ITC was examined using the earlier two-sample task, it was suggested that a ceiling effect may have occurred, making it difficult to establish whether decreasing the probability of ITC had effected discriminability. As the aim here was to develop a titration to avoid such effects, the pre-set criterion level of accuracy used for the titration here was 90% not 100% correct. In the previous experiments here accuracy with the shortest delay in the DMTS task was typically lower than the criterion of 100% correct reached at the end of the titration. This is probably not unexpected, given that accuracy could not be higher than the 100% criterion. Here accuracy with the shortest delay was higher than the level of accuracy attained at the end of the titration. This suggests that for the present task the criterion level of accuracy used to end the titration should be less than 90% correct to allow for any increases in accuracy which may occur on the DMTS task.

In previous experiments it was typically found that accuracy, *a*, was higher on corresponding trials than on non-corresponding trials. This was not generally true for the present experiment. However, the results of the present study were consistent with the previous finding that there was no consistent effect on *b*, across participants, as a function of whether trial pairs were corresponding or not. The present experiment used only two sample stimuli while most of the previous

experiments here, where the data were analysed as a function of whether or not consecutive trials corresponded, used multiple sample-set sizes. The most comparable data set comes from Experiment 7. There, when the one-sample data were excluded and accuracy was the result of performance only with two sample stimuli, accuracy on corresponding trials was still higher than accuracy on non-corresponding trials. Thus, for Experiment 7, using only the two-sample data to calculate corresponding and non-corresponding trials did not seem to have affected accuracy on these trial types. Thus, it is not clear why the result here, for corresponding and non-corresponding trial types differs from that of previous experiments.

For the previous experiments the data were also analysed to establish how accuracy was affected if the non-matching comparison stimulus on trial *n* had appeared as the sample stimulus on trial *n-1*. It was found that this did not occur very often for any of the participants. For the present experiment the non-matching comparison stimuli, on trial *n*, would have appeared as the sample stimulus on the trial *n-1* often. However, in a traditional two-sample DMTS task, when the non-matching comparison stimulus on trial *n* is the same as the sample stimulus on trial *n-1* the sample stimuli must have been different for trials *n* and *n-1*, thus, this must be a non-corresponding pair of trials. Additionally if the non-matching comparison stimulus on trial *n* is different from the sample stimulus on trial *n - 1* the sample stimuli must have been the same for trials *n* and *n-1*, thus, this must be a corresponding pair of trials. Thus, for the present task analysing for correspondence of sample and comparison stimuli provides the same data as analysing for correspondence of sample stimuli.

The order of the experimental trials in the present experiment was pre-determined by the experimenter and was the same for all of the participants. When there are two sample stimuli, for a DMTS task, the theoretical probability of corresponding to non-corresponding trials is 0.50. However, although the trials were initially selected randomly, the probability of ITC was not controlled in the present experiment and the proportion of trials which corresponded was 0.38 and the proportion of trials which did not correspond was 0.62, for all of the participants. This means that for this experiment there were more

non-corresponding trials than there were corresponding trials. Given the results of Experiment 7 it is not clear whether this probability of ITC would have had an effect on accuracy here, as the effect that the probability of ITC has on accuracy has yet to be clarified.

In summary the present data show that it is possible to successfully use disks as sample stimuli in a traditional two sample DMTS task with human participants across a range of delay intervals. Thus, it should be possible to use this type of task to establish whether decreasing the probability of ITC has a systematic effect on matching accuracy on a DMTS task. The titration task used here gave similar levels of accuracy across participants but accuracy was typically higher for the shortest delay in the DMTS task than it was for the 90% correct criterion, reached at the end of the titration. Thus, it was suggested that the criterion level of accuracy required to complete the titration be lowered.

Experiment 10

The aim of the present experiment was to examine, further, the effect of decreasing the probability of ITC on *a* and *b*. Given the findings of Experiment 7 the criterion level of accuracy used for the titration, here was lowered to reduce the probability of a ceiling effect. As a result of the findings of the previous experiment the criterion level of accuracy was set at 80% correct.

The present experiment used white disks in a traditional two-sample DMTS task, with four levels of ITC. The probability of an ITC of 1.0 was not used here, because it required the use of a single-sample condition which is not possible when using a traditional two-sample DMTS task.

Method

Subjects

Eight graduate level psychology students participated in this experiment. Each participant received a book voucher for participating, irrespective of their performance on the experimental task.

Apparatus

The apparatus used here was the same as that used in Experiment 1.

Stimuli

The stimuli were two filled white disks, presented on a blue background. At the beginning of the experiment there was a difference of two pixels between the two stimuli and the stimuli were 98 and 100 pixels in diameter. The size of the sample disks were adjusted during the initial phase of the experiment, for each of the participants.

Procedure

The general procedure was the same as that used for Experiment 9. The procedure differed in that there were two separate experimental sessions. The initial

part of the first session was a two-sample titration. At the beginning of the titration the sample stimuli used were 98 and 100 pixels in diameter. The titration task used here was the same as that used for Experiment 9 with the exception that the criterion level required to complete the titration was 80% correct across 30 trials. At the completion of the titration the DMTS phase of the first session began.

The general procedure used for the DMTS task was the same as that used for Experiment 9 with the exception that here the probability of ITC was varied. The sample stimuli used were the same size as at the completion of the titration phase. There were a total of 340 trials, which were divided into blocks of 17 trials. The first trial in each block of 17 trials was not used in the data analysis as it was a dummy trial which was required to ensure that there were the correct number of corresponding and non-corresponding pairs of trials for each of the conditions. At the completion of each block of 17 trials there was a 30 s inter-block-interval (IBI). During the IBI the screen remained blank and was blue in colour.

The probability of ITC was varied by changing the number of consecutive trials which had the same sample stimuli across a set number of trials in a manner similar to that used for Experiment 9. The probabilities used here were, 0.5, 0.25, 0.125 and 0.0625, and the number of corresponding pairs of trials required for each probability were 8, 4, 2 and 1 out of 16 trial pairs, respectively. The sequence in which the two sample stimuli appeared for each block of trials for each probability are shown in Appendix B. During the first experimental session the first (dummy) trial was always the smaller of the two sample stimuli.

In Condition 1, for the first DMTS phase, there were four blocks of 17 trials. For each of the four blocks of trials there were two delay intervals which were presented for all 17 trials in each block in the following order; 16, 0.05, 0.05 and 16s. There were two probabilities of ITC, used for Condition 1, and each probability was presented once with each of the two delay intervals in the following order 0.5, 0.625, 0.5 and 0.0625. The order in which the delays and probabilities of ITC appeared in Condition 1 is shown in Table 10.1.

For Condition 2 there were four delay intervals. The delay intervals always appeared in the same pseudo-random order as follows; 4, 16, 8 and 0.05 s. As for the

Table 10.1 *The order in which each of the delay intervals and probabilities of ITC appeared in for Condition two, for both the first and the second experimental session.*

Condition	Delay	Probability ITC
	16	0.5
1	0.05	0.0625
	0.05	0.5
	16	0.0625
	4	0.5
	16	0.0625
	8	0.25
	0.05	0.125
	4	0.0625
	16	0.25
	8	0.125
2	0.05	0.5
	4	0.25
	16	0.125
	8	0.5
	0.05	0.0625
	4	0.125
	16	0.5
	8	0.0625
	0.05	0.25

first four blocks of the DMTS phase the trials were arranged into blocks of 17 and the same delay interval was used for all 17 trials in one block. In Condition 2 there were four different probabilities of ITC, 0.5, 0.25, 0.125 and 0.0625. Each of these probabilities appeared with each of the delay intervals, over the 17 trials in the block. The order in which the probabilities appeared with each delay is also shown in Table 10.1. The first experimental session ended when the participant had completed 340 trials at the end of the first DMTS phase. At the completion of the first experimental session the participants were able to take a break before starting the second session of the experiment.

The general procedure of the second experimental session was the same as the first session with the exception that there was no titration procedure. For each participant the samples used for the DMTS task during the second session were the same as those used during the DMTS task in the first session. During the second experimental session the first (dummy) trial, in a block of 17 trials was always the larger of the two stimuli. The second DMTS phase also differed from the first in that the order in which the conditions appeared was reversed, in that Condition 2 was run first and Condition 1 was run second. The second session ended when the participant had completed 340 trials. Once the second experimental session was completed the participants were given information about the purpose of the experiment.

Results

S1 to S8 completed the titration in 50, 140, 70, 60, 40, 50, 70 and 60 trials respectively. At the completion of the titration procedure there was a difference of 6, 20, 10, 8, 4, 6, 10, and 8 pixels in diameter between the two sample stimuli, for S1 to S8 respectively. Thus, S5 entered the DMTS phase of the session with the smallest physical disparity and S2 entered the entered the DMTS phase of the session with the largest physical disparity, relative to the other participants. S1 to S8 achieved 90, 97, 83, 90, 80 83, 97 and 97% correct across the last 30 trials of the titration procedure, respectively.

Figure 10.1a shows percent correct for each of the eight participants and Figure 10.1b shows the average of these, for each of the probabilities of ITC as a

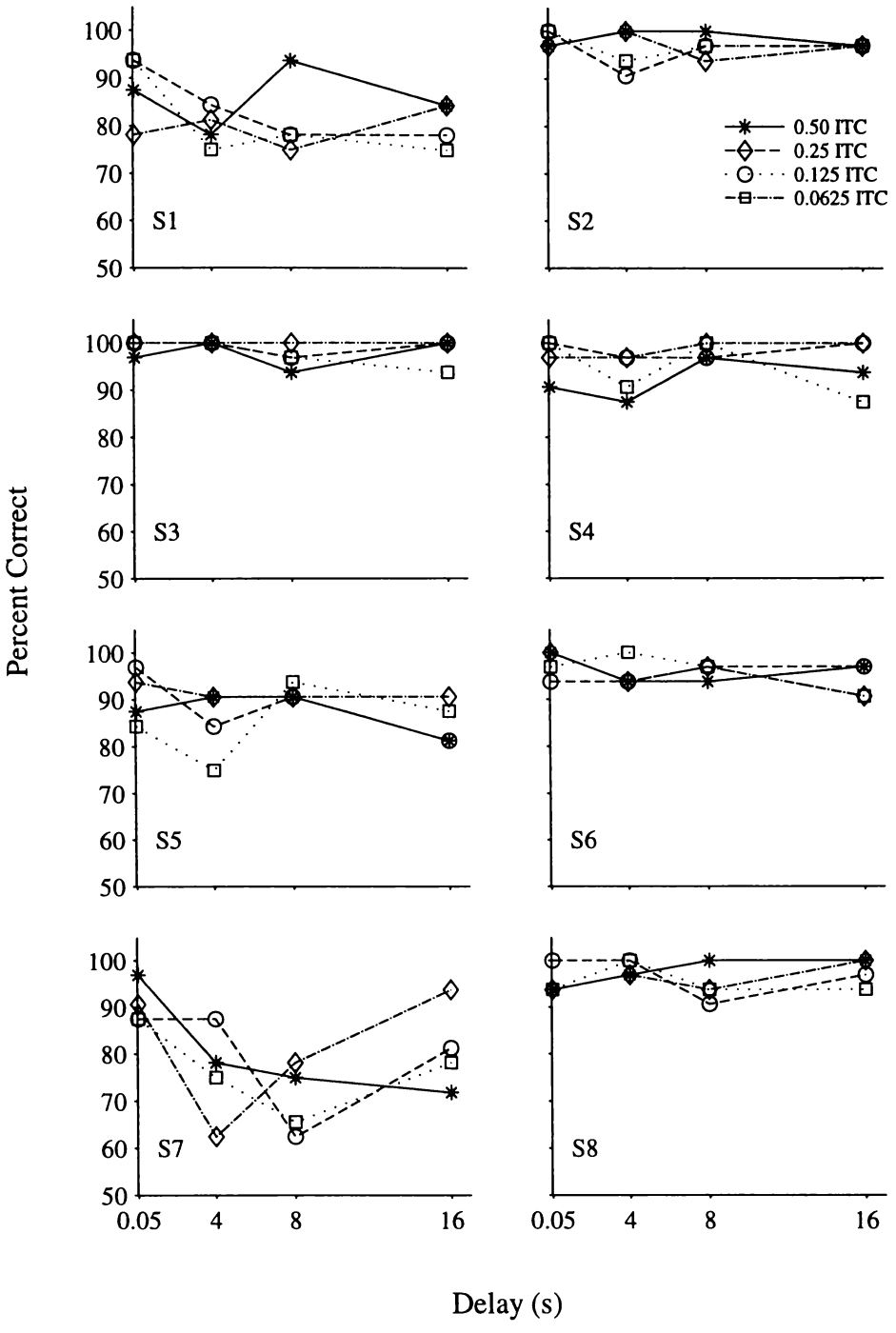

Figure 10.1a. Percent correct for each participant plotted as a function of delay for each probability of intertrial correspondence.

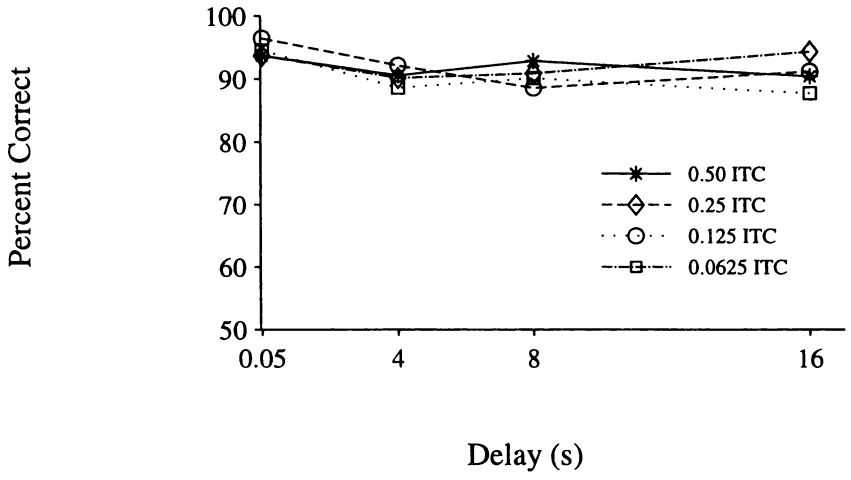

Figure 10.1b. Percent correct averaged across all eight participants plotted for each delay and the probability of intertrial correspondence.

function of delay. Matching accuracy generally varied as the probability of ITC was decreased from 0.50 to 0.0625 but these changes were not systematic for any of the participants nor were they consistent across participants. For participants S2, S3 and S8 accuracy was generally very high and there was little change in matching accuracy as the probability of ITC was decreased. There was generally a slight decrease in matching accuracy as the delay interval was increased, for all probabilities of ITC for most participants. For S7 there was generally a large decrease in accuracy as the delay was increased. The average data are representative of the individual data in that matching accuracy was high for all conditions. As for the individual data matching accuracy changed as the probability of ITC was varied but accuracy did not decrease systematically as the probability of ITC was decreased from 0.5 to 0.0625. The averaged data showed a slight decrease in matching accuracy as the delay interval was increased.

Figure 10.2a shows log *d*, calculated using Equation 6 (Appendix A), for each of the eight participants and Figure 10.2b shows log *d* for the average of these, for each probability of ITC as a function of delay. The individual log *d* data show the same patterns as the percent correct data. There was no consistent change in log *d* at zero delay, as the probability of ITC was decreased from 0.5 to 0.0625. The averaged data are generally representative of the individual data in that log *d* at 0.05 s was high for all conditions and there was little decrement in accuracy as the delay interval was increased.

A repeated measures ANOVA was carried out on these data. Prior to the ANOVA Mauchly's test of sphericity was used to test for homogeneity of variance. It was found that this was not significant for probability of ITC ($W = 0.18$) or delay ($W = 0.59$) but was significant for the interaction ($W = 0.00$) ($\alpha < 0.05$). Given this the Huynh-Feldt (1970) correction was used when testing for significance in the ANOVA. For the ANOVA the main effects of the probability of ITC ($F = 1.54$ $df = 1.97, 13.77$) delay ($F = 2.48$, $df = 3.00, 21.00$) and the interaction ($F = 1.29$, $df = 9.00, 63.00$) were not significant ($\alpha > 0.05$).

Negative exponential functions (Equation 2) were fitted to both the individual and the averaged log *d* data using non-linear regression. Table 10.2 shows the values

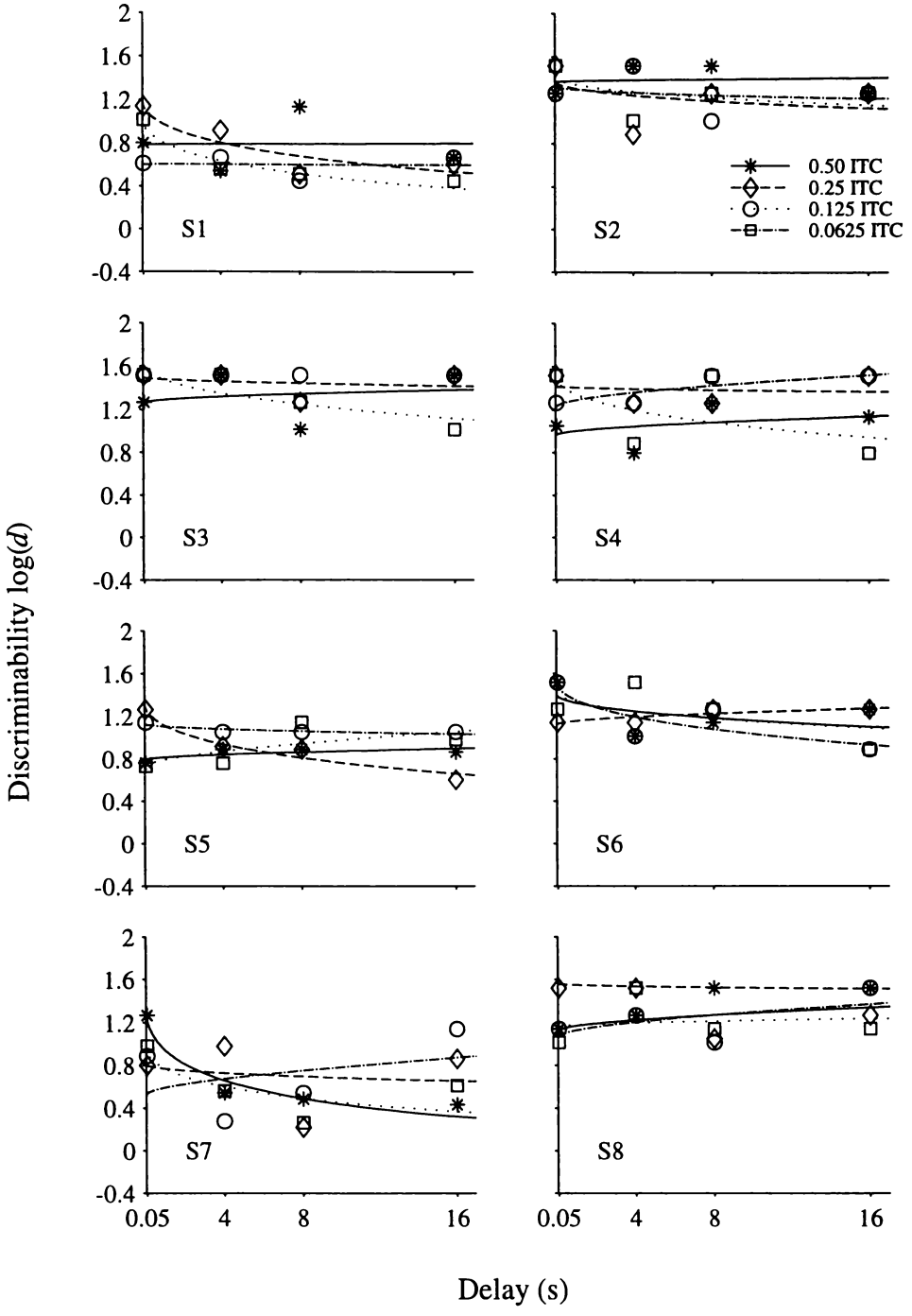

Figure 10.2a. log *d* for the Individual participants plotted as a function of delay and the probability of intertrial correspondence. The fitted functions are negative exponentials (Equation 2).

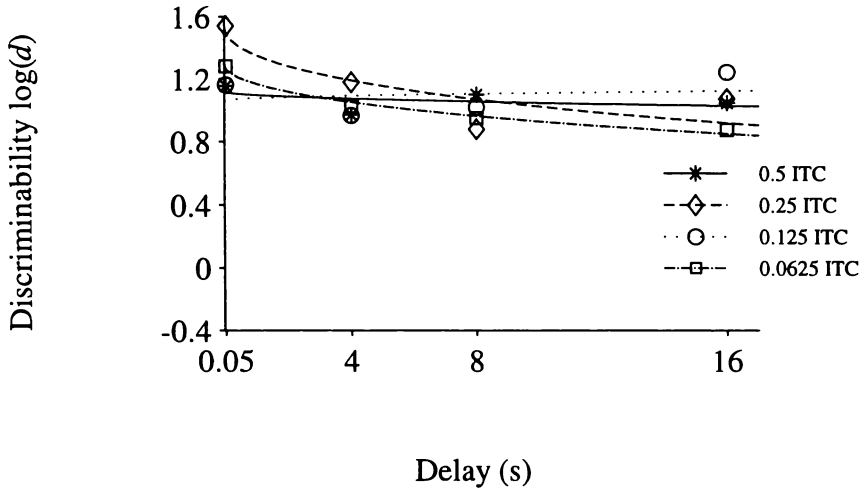

Figure 10.2b. log *d* averaged across all eight participants plotted for each delay and the probability of intertrial correspondence. The fitted functions are negative exponentials (Eqaution 2).

Table 10.2 *a, b (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions for each of the experimental conditions.*

	Probability of ITC	*a*	*b*	VAC %	SE
	0.5	0.78	-0.01	0	0.12
S1	0.25	1.20	0.20	81	0.11
	0.125	0.59	0.00	0	0.09
	0.0625	1.04	0.25	94	0.05
	0.5	1.37	-0.01	1	0.13
S2	0.25	1.37	0.05	13	0.21
	0.125	1.32	0.02	4	0.17
	0.0625	1.41	0.05	23	0.16
	0.5	1.26	-0.02	4	0.20
S3	0.25	1.50	0.01	6	0.11
	0.125	1.52	*	*	*
	0.0625	1.63	0.09	74	0.11
	0.5	0.96	-0.04	14	0.16
S4	0.25	1.42	-0.01	1	0.13
	0.125	1.22	-0.06	71	0.07
	0.0625	1.52	0.12	33	0.28
	0.5	0.79	-0.03	53	0.04
S5	0.25	1.33	0.17	96	0.05
	0.125	1.13	0.02	74	0.02
	0.0625	0.72	-0.10	50	0.12
	0.5	1.41	0.06	29	0.16

S6	0.25	1.12	-0.03	71	0.03
	0.125	1.53	0.12	70	0.13
	0.0625	1.43	0.07	30	0.18
	0.5	1.34	0.36	94	0.08
S7	0.25	0.81	0.05	3	0.29
	0.125	0.52	-0.13	8	0.32
	0.0625	0.98	0.24	56	0.17
	0.5	1.12	-0.08	86	0.06
S8	0.25	1.56	0.07	43	0.46
	0.125	1.06	-0.06	31	0.15
	0.0625	1.16	0.02	2	0.19
	0.5	1.12	0.02	20	0.06
Average	0.25	1.54	0.13	73	0.13
	0.125	1.06	-0.02	4	0.12
	0.0625	1.30	0.11	98	0.02

* Indicates that the negative exponential function could not be fitted to the data for these sample-set sizes.

obtained for the parameters *a* and *b* for the individual data and the averaged data as well as the standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions. For S3, S4 and S6 *a* tended to increase as the probability of ITC was decreased. There was generally no systematic change in *a* as the probability of ITC was decreased for the remaining 5 participants or the averaged data. A bi-directional Ferguson's test for trend (Ferguson, 1965) showed that *a* did not trend significantly as the probability of ITC was decreased ($\alpha > 0.05$). A bi-directional Ferguson's test for trend (Ferguson, 1965) also showed that *b* did not trend significantly as the probability of ITC was decreased ($\alpha > 0.05$).

Figure 10.3 shows percent correct for each participant for each probability of ITC used in Condition 1 in the first and the second experimental sessions, for both delay intervals. In general there was no consistent change in matching accuracy for any of the participants across the experimental sessions for either probability of ITC with either delay. However, for S4 and S8 matching accuracy for the probability of ITC of 0.5 was higher in the second session than in the first session, for both delays. For S5 and S7 matching accuracy for the probability of ITC of 0.0625 was higher in the second session than in the first session, for both delays. For S6 matching accuracy for the probability of ITC of 0.5 was higher for the second session than for the first session and matching accuracy for the probability of ITC of 0.0625 was higher for the first session than for the second session, for both delays.

Figure 10.4a shows log *d*, for corresponding and non-corresponding trials, for each of the participants, and Figure 10.4b shows log *d* for the average of these as a function of delay interval for each level of ITC concatenated across all experimental conditions and both sessions. For five of the individuals and for the averaged data log *d* for both the corresponding and the non-corresponding trial pairs, was high for all delays, the exceptions being S1, S5 and S7. There was no consistent change in *a* as a function of whether trials were corresponding or non-corresponding. For the averaged data the *a* values for the corresponding were slightly higher for all delay intervals than for non-corresponding trial pairs.

Negative exponential functions were fitted to the log *d* data for the corresponding and non-corresponding trial pairs, using non-linear regression.

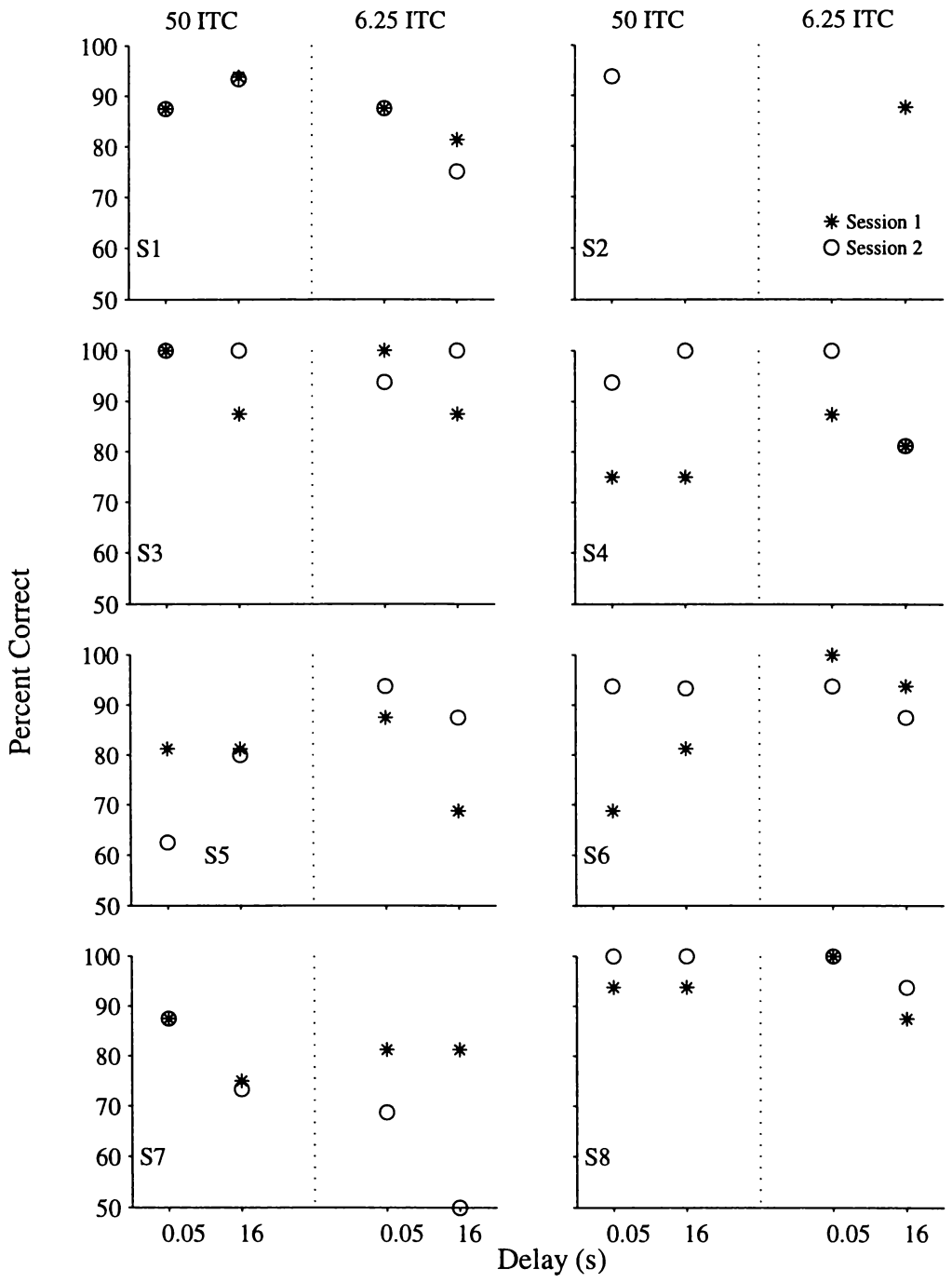

Figure 10.3. Percent correct for each participant plotted as a function of delay for each probability of intertrial correspondence for Condition 1 of the first and the second experimental sessions.

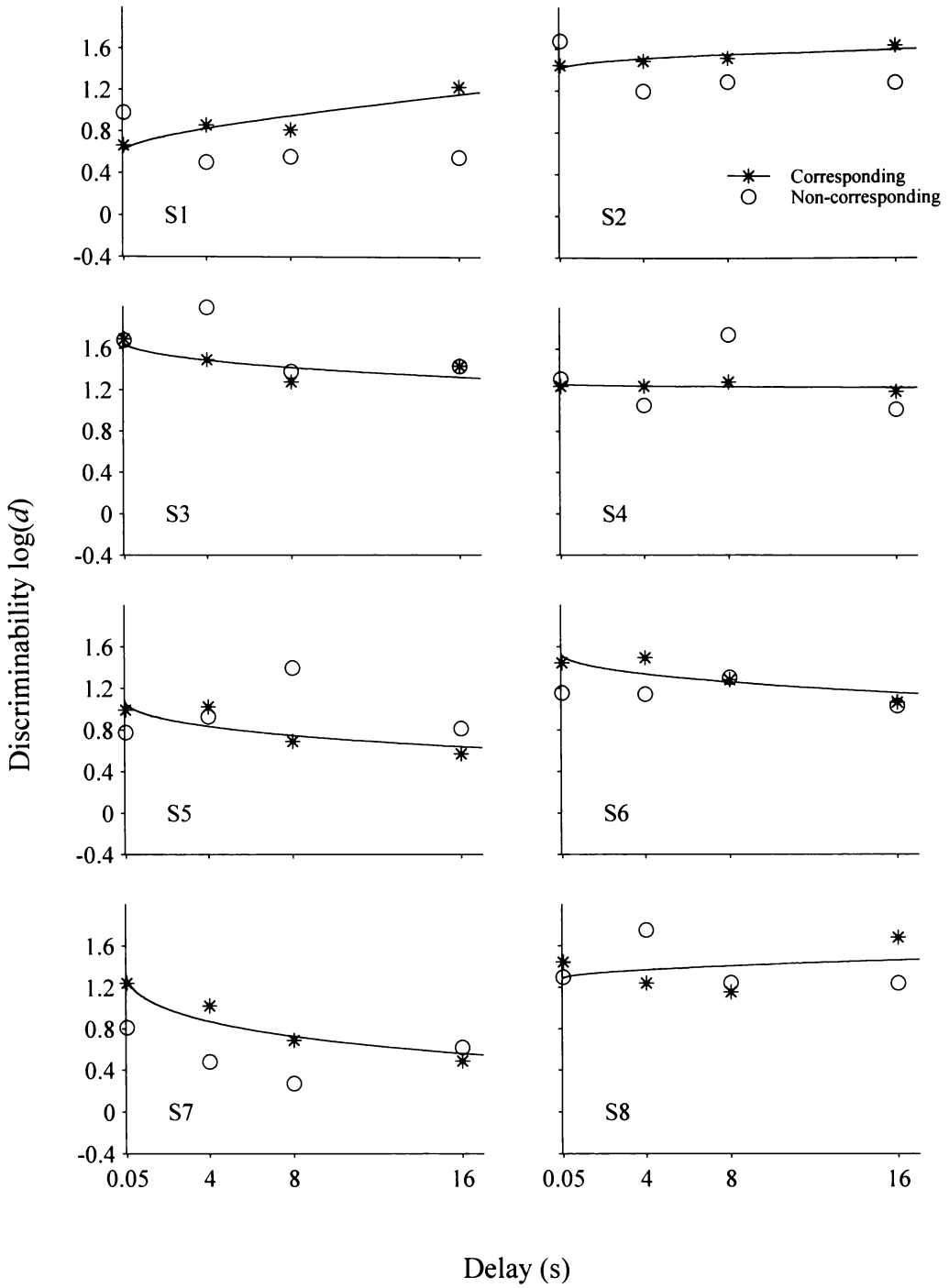

Figure 10.4a. log *d* for each participant plotted as a function of delay for corresponding and non-corresponding trials. The fitted fuctions are negative exponentials (Equation 2).

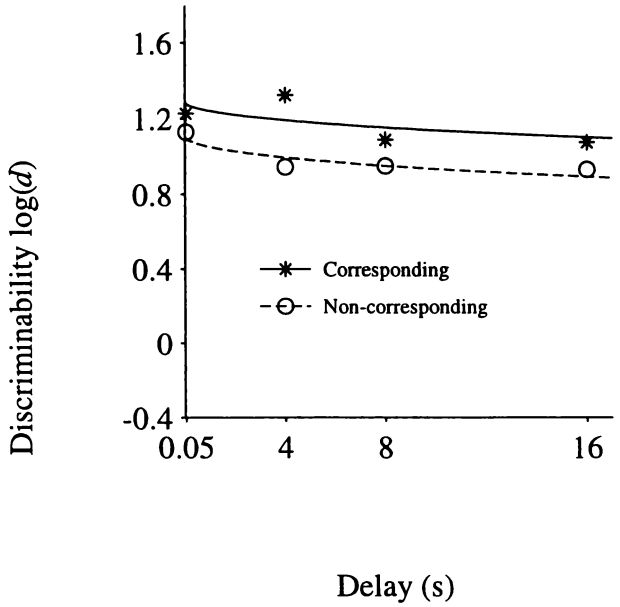

Figure 10.4b. log *d* averaged across eight participants plotted as a function of delay for corresponding and non-corresponding trials. The fitted functions are negative exponentials (Equation 2).

Table 10.3 shows the values obtained for the parameters *a* and *b* for the individual and the averaged data as well as the standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions. For S1, S2, S3, S4 and S8, *a* was higher when trial pairs did not correspond than when they did. For S5, S6 and S7, *a* was higher when trial pairs corresponded than when they did not. For the average data *a* values were higher for corresponding trials than for non-corresponding trials. Accuracy generally decreased slightly as the delay interval was increased for most participants, as illustrated by the small *b* values in Table 10.3. For some participants accuracy increased as the delay interval was increased, as shown by the negative *b* values in Table 10.3. In all cases these increases in accuracy, as the delay interval was increased, occurred for corresponding trials. There was no systematic difference in *b* as a function of ITC for the individual, with the exception of S7 or the averaged data. The percentage of VAC was lower for non-corresponding trials than for corresponding trials for all of the participants. The standard error of estimate was generally low for all participants but for all participants it was lower for corresponding than for non-corresponding data.

Log *b*, response bias was calculated (Equation 8, Appendix A) for the individual participants and for the averaged data based on the size of the comparison stimuli, for each probability of ITC and delay. The individual data are shown in Figure 10.5a and the averaged data are shown in Figure 10.5b. There was generally little or no bias towards either the smaller or the larger of the two comparison stimuli for S3. The participants, S4, S5 and S7 showed a tendency towards responding to the smaller of the two comparison stimuli (values > 0). S2 and S8 showed some bias towards responding to the larger of the two comparison stimuli (values < 0). For S1 and S6 any bias tended to be variable. None of the participants showed a tendency for this bias to change systematically with probability of ITC or with delay. The averaged data show that there was an overall tendency to respond to the smaller of the two comparison stimuli but there was no systematic change in this bias as a function of the probability of ITC or as a function of delay.

Log *b* response bias was also calculated (Equation 8, Appendix A) for the individual participants and for the averaged data based on the side of the screen on

Table 10.3 *Estimates of a, b, (Equation 2) standard error of estimate and the percentage of variance accounted for (VAC) by the fitted functions for corresponding and non-corresponding trials for each of the participants and the average of these.*

	Trial Types	*a*	*b*	VAC %	Std Err
S1	CT	0.59	-0.17	84	0.08
	NCT	0.96	0.19	74	0.10
S2	CT	1.40	-0.04	86	0.03
	NCT	1.61	0.07	56	0.12
S3	CT	1.67	0.06	62	0.09
	NCT	1.82	0.05	25	0.21
S4	CT	1.26	0.01	14	0.03
	NCT	1.34	0.02	2	0.29
S5	CT	1.09	0.13	68	0.11
	NCT	0.88	-0.04	6	0.24
S6	CT	1.54	0.07	65	0.10
	NCT	1.19	0.01	5	0.09
S7	CT	1.35	0.22	91	0.09
	NCT	0.77	0.17	33	0.16
S8	CT	1.27	-0.03	8	0.19
	NCT	1.45	0.02	4	0.21
Average	CT	1.11	0.05	81	0.04
	NCT	1.29	0.04	40	0.08

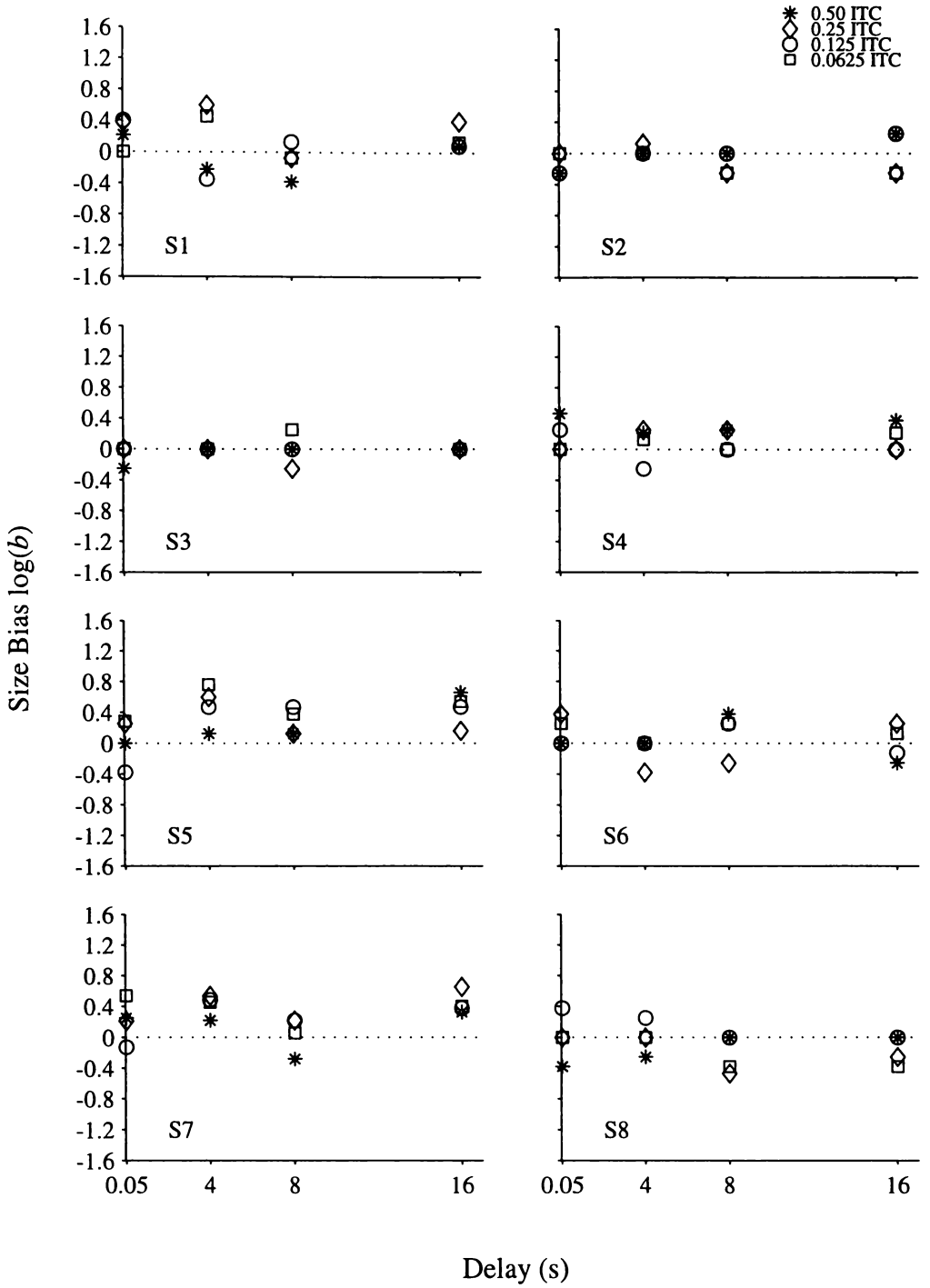

Figure 10.5a. Response bias towards the size of the comparison stimuli (log *b*) for the individual participants plotted as a function of delay and the probability of intertrial correspondence.

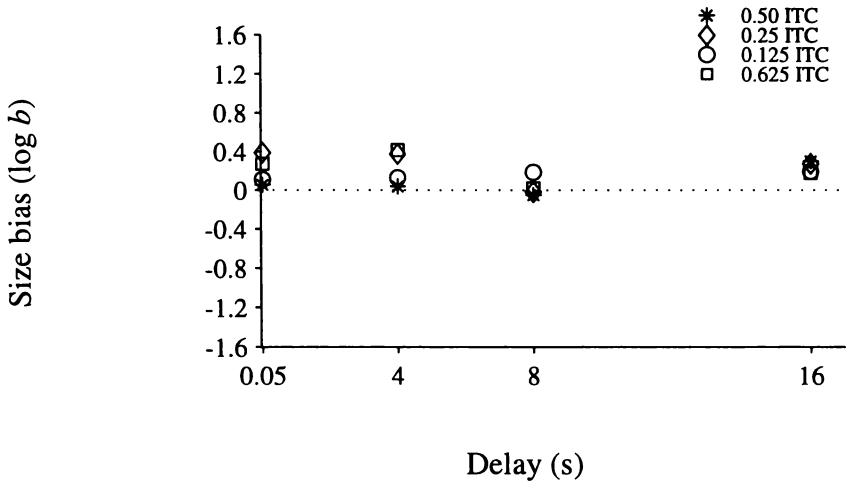

Figure 10.5b. log *b* as a function of the size of the comparison stimuli averaged across all eight participants plotted for each delay and the probability of intertrial correspondence.

which the comparison stimuli appeared, as a function of ITC and delay. The individual data are shown in Figure 10.6a and the averaged data are shown in Figure 10.6b. Four of the participants (S1, S2, S5 and S6) showed a tendency for bias to shift across left and right responses. S7 and S8 showed a bias towards choosing the comparison stimulus on the left of the screen, while S4 showed a bias towards choosing the comparison stimulus on the right of the screen. There was no systematic change in this bias as a function of the probability of ITC or delay. The averaged data show that there was generally little or no bias towards responding to the left or the right alternative as a function of the probability of ITC or delay. There was no tendency for bias to change as a function of the probability of ITC or delay.

Discussion

In the present study accuracy decreased only slightly as the delay interval was increased for most participants, as shown by the generally small *b* values for the fitted functions. There were no systematic changes in *b* as the probability of ITC was decreased. Accuracy varied as the probability of ITC was changed but these changes in accuracy did not occur systematically as the probability was decreased, nor were they consistent across participants. This finding is consistent with that of Experiment 7.

There was a consistent tendency here, for some of the participant's responding to be biased towards selecting either the smaller or the larger of the two comparison stimuli. This finding is consistent with that of Experiment 5, where the number of sample-stimuli varied from 1 to 32, and for Experiments 8 and 9, where a traditional 2-sample DMTS task was used. In addition to this, it was also found here that there was a systematic tendency for participant's responding to be biased towards selecting either the comparison stimulus on the left or on the right of the screen. This finding was not consistent with the findings of the previous experiments.

White (1985) points out that logit *p* is equivalent to log *d* in cases where there is no response bias (log *b*). Thus, when there is no bias it should be possible to compare data analysed with logit *p* to data analysed with log *d*. For most of the

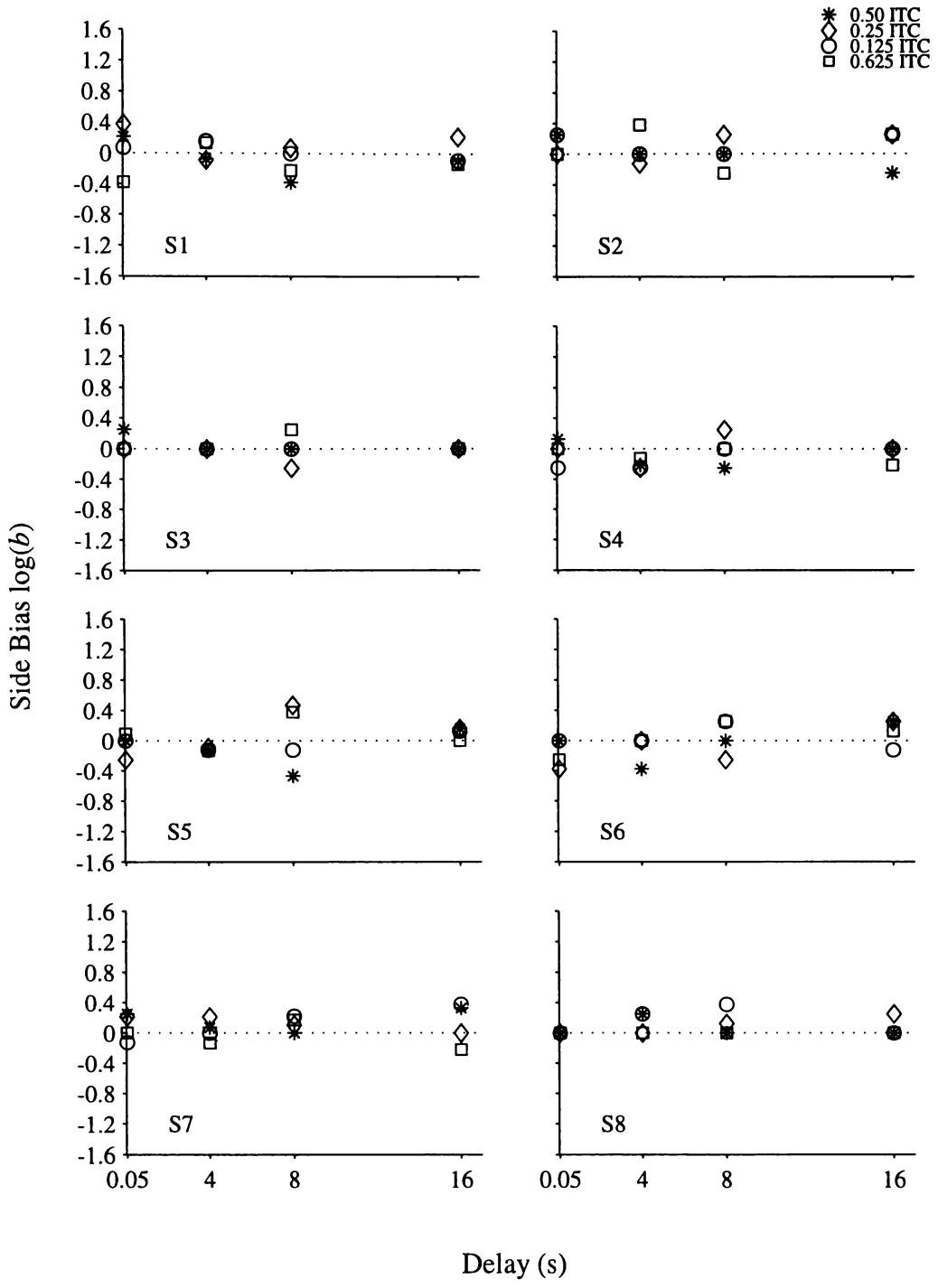

Figure 10.6a. Response bias towards the side of the screen on which the comparison stimulus was presented (log *b*) for the Individual participants plotted as a function of delay and the probability of intertrial correspondence.

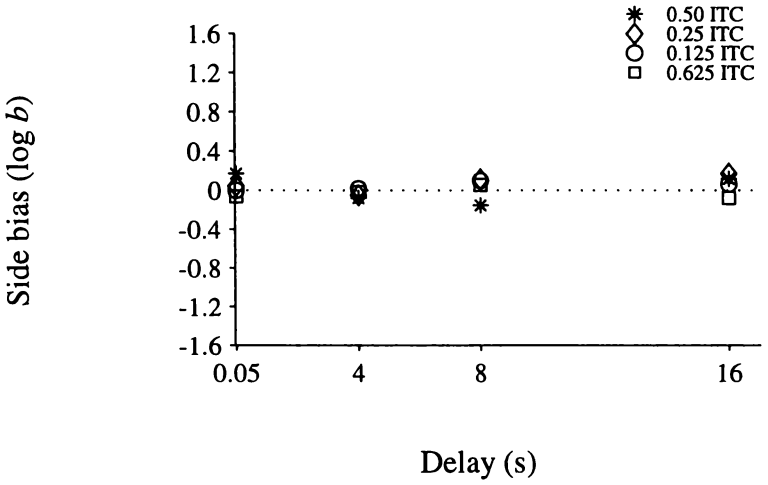

Figure 10.6b. log *b* as a function of the side of the screen on which the comparison stimuli appeared, averaged across all eight participants, plotted for each delay and the probability of intertrial correspondence.

experiments conducted here (Experiment 1 to Experiment 7) it was not possible to calculate log *b* and thus, to know whether logit *p* and log *d* were equivalent. However, it was possible to calculate an estimate of response bias (log *b*) for Experiments 1 to 7. As mentioned above, for all of these experiments, with the exception of Experiment 5, this estimate of response bias showed that participants' data had little or no bias, either towards choosing a particular sized alternative or towards choosing an alternative displayed on a particular side of the screen. Thus it might be expected, if these estimates of log *b* were accurate that for these experiments, with the exception of Experiment 5, that logit *p* provided a good estimate of log *d*.

For Experiments 8 to 10, where a traditional two-sample DMTS task was used, it was possible to calculate log *d* and log *b*. It is also is also possible to calculate logit *p* and to estimate response bias, as the logarithm of the ratio of the total number of times alternative 1 was selected to the total number of times alternative 2 was selected, for these data. Given this, data from Experiment 10 are reanalysed here and comparisons are made between logit *p* and log *d* and the estimate of bias, described above, and log *d*.

Figure 10.7 shows log *b* compared to the estimate of bias, described above, plotted for each of the experimental conditions for each of the delays, averaged across all participants. This analysis shows that the estimate of bias used for the earlier experiments tended to underestimate the level of response bias, log *b*. Thus, although the estimate gives the direction in which response bias moves, for a given participant, it does not generally provide a good estimate of log *b*. Thus, bias may have been underestimated in Experiments four to seven.

Figure 10.8 shows a comparison of the difference between log *d* and logit *p* (Difference = log *d* – logit *p*), as a function of log *b*, averaged across all participants and across all conditions and delays. It shows that when there was response bias, as measured by log *b*, logit *p* differed from log *d* and that as log *b* increased the difference between logit *p* and log *d* increased. This confirms that logit *p* does not provide a good estimate of log *d* when there is bias. This increase in the difference between log *d* and logit *p* as response bias was increased also suggests that data

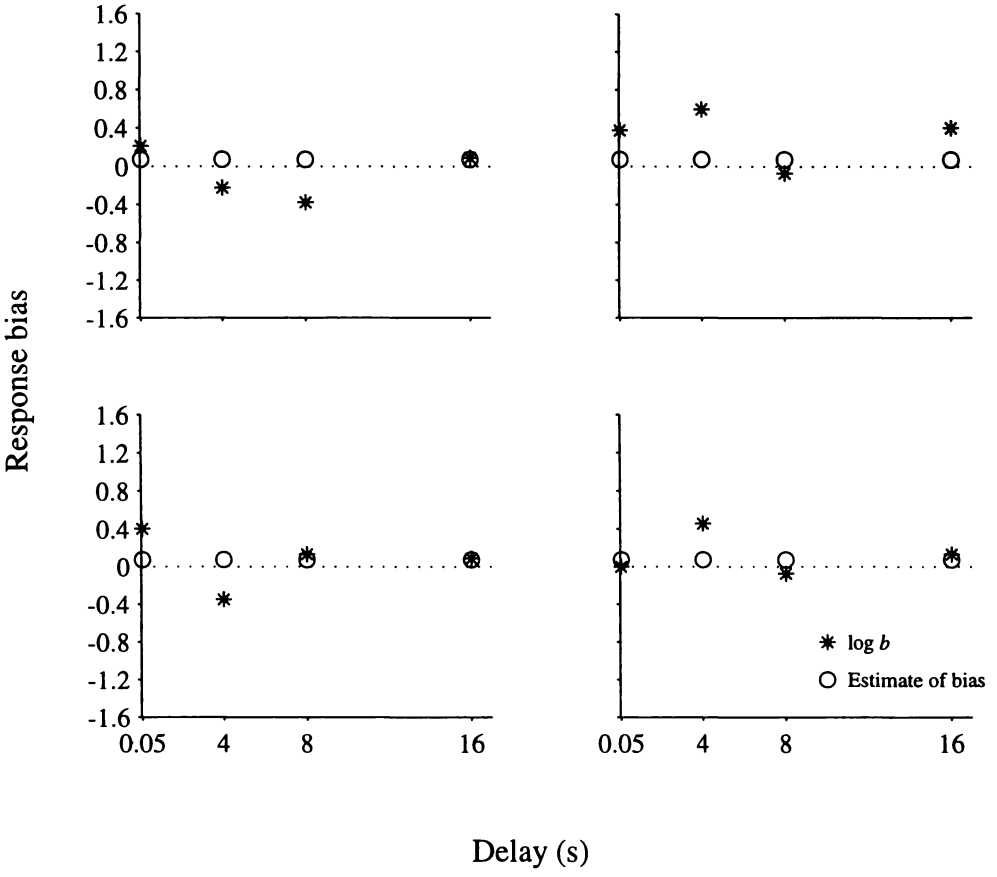

Figure 10.7. Log *b* and an estimate of response bias, averaged across all participants, plotted for each of the experimental conditions and for each of the delays.

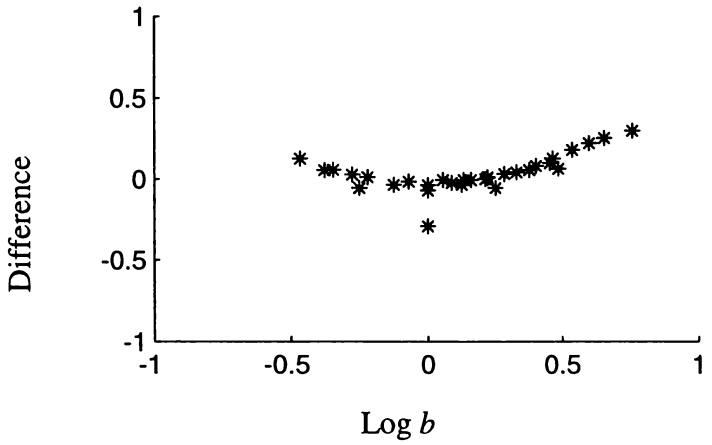

Figure 10.8. The difference between log d and logit p, averaged across all participants, plotted as a function of log b.

should not be compared across experiments when one set of data has been analysed using log *d* and the other using logit *p* unless there is clearly no response bias.

In the present experiment accuracy was very high for three of the participants and varied little as the probability of ITC was decreased and in this respect the findings here are very similar to those of Experiment 7. Thus, for these participants the present procedure also seems to have resulted in a ceiling effect, as in Experiment 7. However, for the remaining five participants there was no evidence of a ceiling effect and matching accuracy still did not change systematically as the probability of ITC was decreased. Thus, the present data provide support for the earlier finding in Experiment 7, that when the sample-set size is held constant matching accuracy does not decrease systematically as the probability of ITC decreases.

In the previous experiments when the sample-set size was varied it was typically found that initial discriminability, *a*, was higher on corresponding than on non-corresponding trials. However, when responding on corresponding and non-corresponding trials was concatenated across all experimental conditions and compared here, it was found that that this was not consistently true for all participants. Thus, the findings here are consistent with those of Experiment 9.

The present experiment used a traditional two-sample DMTS task similar to that used by Edhouse and White (1988a) to investigate the impact of ITC on matching accuracy. They found that when performance on corresponding trials was compared to performance on non-corresponding trials that *b* varied but *a* did not. However, for the present study there was no consistent change in *a* or *b* as a function of whether or not trials were corresponding or non-corresponding. Thus, the finding here is not consistent with either the findings of the previous multiple-sample experiments or with the findings of Edhouse and White's (1988a). It is, however, consistent with that of the previous experiment, where a similar traditional two sample DMTS task was used. It is not clear why these differences occurred.

The present experiment suggests that the decrease in *a* which occurred as the number of sample stimuli was increased in previous experiments was probably not due to the co-varying decrease in the probability of ITC. Thus, it seems unlikely that the decrease in the probability of ITC which occurred as the number of sample

stimuli was increased would account for an orderly decrease in accuracy. This suggests that the previous decrement in accuracy, found when the number of sample stimuli was varied, was due to some other effect of increasing the sample-set size.

General Discussion

As previously discussed, studies which examined the affect that sample-set size has on matching accuracy, with a DMTS task, gave three inconsistent findings. One, that matching accuracy did not change as the sample-set size was increased (Etkin and D'Amato, 1969; Mason and Wilson, 1974). Two, that matching accuracy increased as sample-set size was increased (Worsham, 1975; Mishkin and Delacour, 1975). Three, that matching accuracy decreased as sample-set size was increased (Roberts, 1980; Adamson, 1995). The present thesis investigated parametrically the affect that increasing sample-set size has on matching accuracy with human participants for a DMTS task. It was found that increasing the sample-set size (Experiments 3, 5 and 6) resulted in a systematic decrease in matching accuracy. When a negative exponential (Equation 2) was used to describe these data it was typically found that increasing the sample-set size resulted in a systematic decrease in initial discriminability, a, but did not typically have a systematic effect on the rate of decay, b. Thus the present result provided support for Roberts (1980) and Adamson (1995).

It was suggested earlier that the decrease in a, as the sample-set size increased, may have been due to a decrease in the probability that sample stimuli on consecutive trials were the same. That is, as sample-set size increased the probability of ITC decreased. This suggestion was investigated using a DMTS task, similar to that used for the experiments discussed above, where the probability of ITC was varied and the number of sample stimuli was held constant (Experiment 7). It was expected that initial discriminability, a, would decrease as the probability of ITC was decreased. However, decreasing the probability of ITC did not affect either a or b. It was suggested that these results were confounded by a ceiling effect and so it was not clear whether the result was a true indication of how accuracy would be affected as the probability of ITC was decreased.

It was decided that a more traditional two-sample procedure, similar to that used by Edhouse and White (1988a), would be used to investigate the impact of varying ITC on matching accuracy further. This type of DMTS task is typically

used in discrimination tasks with animal subjects, however, the present study (Experiments 8, 9 and 10) showed that it could also be successfully used with human participants. The present study (Experiment 10) also showed that when there was no ceiling effect that decreasing the probability of ITC, or the ratio of corresponding to non-corresponding trial pairs, did not result in a systematic change in discriminability *a* or in the rate of forgetting, *b*. This finding is consistent with that of Experiment 7 and suggests that the decrease in *a* which occurred as the sample-set size was increased was not due to the decrease in the probability of ITC.

As mentioned in the introduction two-sample DMTS tasks have been used frequently to study memory with animals but have not typically been used with human participants. Instead studies of human memory, which have used a DMTS task, have often used more than two sample stimuli (Parr (1992) used eight samples, Money, Kirk and McNaughton (1992) used thirteen, Adamson (1995) used four and Holdstock et al. (1995) used trial unique stimuli). Given that the number of sample stimuli used with humans often varies across studies any differences in the number of sample stimuli should also be taken into account when comparing findings across human studies. For example, if two studies, with humans, were attempting to investigate the impact of increasing stimulus discriminability on matching accuracy and all aspects of the experimental procedure were the same except for the sample-set size used the findings may not be directly comparable.

However, it has been shown here (Experiments 9 and 10) that it is possible to use a traditional two-sample DMTS task with human participants successfully. The use of such a two-sample DMTS task has two main advantages. The first is that it would be possible to calculate log *d* and thus, it would not be necessary to use logit *p* to estimate log *d*. As shown earlier logit *p* only provides a good estimate of log *d* when there is little or no bias. A second advantage of using a two-sample DMTS task with humans is that the findings would be more directly comparable to animal studies which have used such a two-sample task.

There is possibly an alternative way to use a two-sample DMTS task and to vary the number of sample stimuli. This most easily illustrated by an example. There could be two pairs of stimuli, A and B and C and D, where A and B would appear only as comparison stimuli on trials where A or B were sample stimuli and C and D would only appear as comparison stimuli on trials where C or D were sample stimuli. Thus, it would be possible to calculate log *d* separately for trials with each pair of samples. Accuracy, log *d*, for A and B trials from this four-sample condition could then be compared to accuracy from a condition with A and B trials only. Adding further pairs of stimuli would show whether increasing the sample number had an affect on matching accuracy as it would be possible to compare accuracy on the different trial types over several conditions, using log *d*. It remains to be seen whether such a procedure would be successful.

This thesis also developed and successfully used a titration task to equate initial levels of accuracy across participants, both with a traditional two-sample and a multiple sample DMTS task. The use of such a titration task has two main advantages. The first is that reducing inter-participant variability, in *a*, means that it is not necessary to account for differences in different participants ability to do a task when interpreting the results of a task. That is, using a titration task to equate performance across participants allows for the effects of the experimental conditions to be interpreted independently of different participants ability to do the task. The second is that it is possible to manipulate the initial level of accuracy, *a*. This avoids ceiling and floor effects in the data which may mask any effect of the experimental conditions. More specifically this means that if accuracy was expected to decrease across a range of experimental conditions, the titration task could be used to ensure that initial accuracy was high, so that there was room for accuracy to decrease. Alternatively if accuracy was expected to increase, the titration task could be used to ensure that initial accuracy was low, so that there was room for accuracy to increase. For example, here where it was possible that feedback might result in increased accuracy (Experiment 4) the criterion level of accuracy required for the titration was low. For a task such as the traditional two-sample DMTS where the difficulty of the task cannot be manipulated by

increasing the number of sample stimuli the titration is particularly valuable as it allows the difficultly of the task to be determined for each individual, avoiding the task being to easy for some participants and to hard for other participants.

Edhouse and White (1988a) state that responding on trial *n* is not necessarily completely under the control of the conditional discrimination on trial *n* and that the response may be influenced by events occurring prior to the current trial. This statement suggests that experimental trials are not independent of each other. If trials were independent, we would not expect one trial to have an impact on the accuracy on another trial. In support of Edhouse and White (1988a) here, there were two clear examples that accuracy on a given trial was affected by previous trials. The first example of this is that accuracy here (Experiments 5, 6 and 7) was affected by whether or not consecutive pairs of trials were corresponding or non-corresponding. As such, this effect is calculated across pairs of consecutive trials and therefore relies upon the assumption that at least trials *n* and *n* – 1 impact on each other. The second example is that here (Experiments 3, 5 and 6) as the number of sample stimuli was increased accuracy tended to decrease. This finding suggests that accuracy is affected by more than just the previous trial. That is, in order for this effect to occur the participant must be exposed to a range of stimuli across a number of trials. These findings suggest that experimental trials are not independent and that this assumption should not be made when conducting such experiments. However, few experimental studies, reviewed by this experimenter, appeared to account for the possibility that trials may not be independent. The suggestion that trials are not independent for DMTS tasks may have some implications for other areas of research. For example, the MTS task is the same as the DMTS task, except for the delay between the presentation of the sample and the comparison stimuli. Thus, if trials with a DMTS task are not independent then it is likely that trials with a MTS task are not independent. In addition, if trials with a discrete trial task such as the DMTS task are not independent then it is possible that trials with other discrete trial tasks, such as those frequently used in psychophysics, are not independent.

For all experiments and for most participants here, it was generally true that changes in accuracy with increasing delay were not large (i.e. *b* was generally small). In addition none of the experimental changes appeared to have had any systematic impact on *b*. Thus, in the case of the present study, *b* was not affected by sample-set size or by the probability of ITC. Although this finding seems clear it is possible that any changes in *b* were masked by the lack of any effect of delay. It is possible to increase the effect of delay on accuracy and hence increase *b*. For example, White (1985) introduced a house-light into the delay interval, in a DMTS task with pigeons, and found that this increased *b*, but not alter *a*. Peterson and Peterson (1959) used a distracter task during the delay interval, reasoning that accuracy could be decreased by preventing participants from verbal repetition or rehearsal of the stimuli during the delay. Such a distracter task might also increase *b*. Increasing *b* and then varying the sample-set size and the probability of ITC might help clarify whether or not it was the small size of *b* that masked any changes in *b* here.

As mentioned earlier, it was found here, that increasing the number of sample stimuli resulted in a decrease in initial accuracy *a*. The suggestion that this was the result of changing ITC was not found to be the case. As the number of sample stimuli was here increased the range over which the sample stimuli varied in size also increased. That is, as the sample-set size was increased the physical disparity between the sample stimuli was held constant, with larger and smaller stimuli added to the set to increase the number of stimuli. This means that each time larger and smaller stimuli were added to the set the overall range of sample sizes was also increased (see Table 5.1 and Table 6.1).

Hinson and Lockhead (1986) and Lockhead and Hinson (1986) suggested that it is not the number of sample stimuli in a set but the range over which stimuli vary which impacts on accuracy. Hinson and Lockhead (1986) and Lockhead and Hinson (1986) used an auditory discrimination task where the number of sample stimuli was held constant and the physical range over which the sample stimuli varied was increased by making stimuli more physically different from each other. It would generally be expected that increasing the physical difference or disparity

between sample stimuli would result in increased accuracy, as suggested by White (1985). However, both of these studies found that increasing the range or spread of stimuli along a physical dimension resulted in decreases in accuracy. They called this finding the range effect. This effect has also been found for other studies (Gravetter & Lockhead, 1973; Ward and Lockhead, 1971).

Given the findings discussed above and that here, as the number of sample stimuli was varied the range also varied data from Experiments 5 and 6 were reanalysed in a manner analogous to that of Lockhead and Hinson (1986). They state that range effects are measured between conditions by comparing the average of all responses to a particular stimulus across small and large ranges.

Here the average performance with each of the four different sized sample stimuli, which occurred in all of the larger sample-sets, was compared across conditions where the sample-set size was varied and hence the size range of stimuli varied. Thus, performance for a particular stimulus was compared across small and large ranges. The stimuli used, for the reanalysis of data here, were from the four-sample condition (84, 87, 90, and 93 pixels). For Experiments 5 and 6 average performance, for each of the sample stimuli, was compared across the four-, the eight-, and the sixteen-sample and for Experiment 5 performance was also compared across the thirty-two-sample condition. The relation between these sample-set sizes and the range of stimulus sizes is illustrated in Table 5.1 and Table 6.1.

For each sample stimulus, for each sample-set size, accuracy was averaged across all delays. Accuracy was averaged across delays, as although there was a decrement in accuracy across delays this decrement did not vary systematically across conditions so any effect of averaging across delays should be consistent across the sample-set sizes. This averaging across delays was also necessary to maintain the number of trials able to be analysed for the larger sample-set sizes where the number of times each of the stimuli appeared as a sample stimulus was much smaller than for the smaller sample-set sizes.

For both Experiment 5 and 6 the average logit p was obtained for each sample stimulus across each of the sample-set sizes and these data are shown in

Figure 11.1 (Experiment 5) and Figure 11.2 (Experiment 6). Figure 11.1 shows that there was a notable and systematic decrease in accuracy for all samples, with the exception of the 93 pixel sample, as the sample-set size, and hence size range, was increased from four to eight and to sixteen. For the 93 pixel sample there was only a very small decrease in accuracy from the four to the eight sample-set size. For all sample stimuli there was a notable increase in accuracy from the 16 to the 32 sample-set size and for all samples, with the exception of the 84 pixel sample, accuracy was highest when there were 32 samples when the range was greatest.

Figure 11.2 shows that for the 87 and the 90 pixel sample stimuli there was a notable decrease in accuracy from the 4- to the 8-sample condition and for the 90 pixel sample there was also a small decrease from the 8- to the 16-sample condition. For the 84 and the 93 pixel samples there were generally only very small decreases in accuracy as the sample-set size and hence the range was increased.

For Experiment 5 the results for the 84, 87 and the 90 pixel samples across the four-, the eight- and the sixteen-sample conditions are consistent with Hinson and Lockhead (1986) and Lockhead and Hinson (1986). However, the large increase in accuracy from the 16- to the 32-sample set is not consistent with their suggestion that as range increases accuracy decreases, nor is it consistent with the earlier suggestion here that as sample-set size increases matching accuracy decreases. It is not clear why this increase occurred.

For Experiment 6 any effect of range or sample-set size on accuracy for each of the sample-set sizes or range was small. Although for the 90 and the 93 pixel samples there was a decrease in accuracy as the number of sample stimuli was increased from four to eight. Thus, the present reanalysis of data does provide some support for the suggestion that increasing the physical range over which stimuli vary results in a decrease in accuracy. However, it should be noted that here, even with the analysis suggested by Lockhead and Hinson (1986) increases in sample-set size and range are still confounded. Therefore, it is not possible to say which of the two factors any effect might be attributable to. As such, it would be necessary to experiment further in order to determine whether it is the number

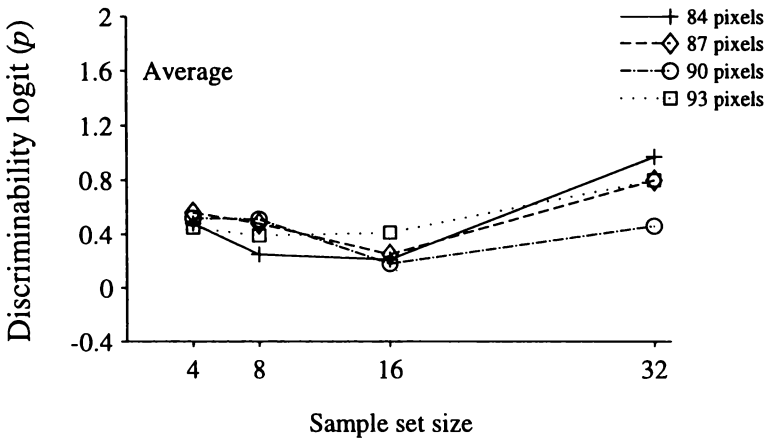

Figure 11.1. Logit *p* for each sample stimulus from the four-sample condition, from Experiment 5, plotted as a function of sample-set size, averaged across participants and delays.

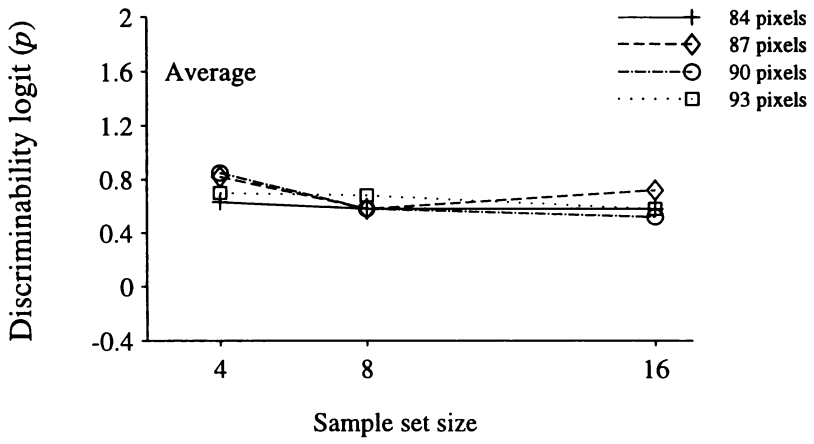

Figure 11.2. Logit *p* for each sample stimulus from the four-sample condition, for Experiment 6, plotted as a function of sample-set size, averaged across participants and delays.

of sample stimuli or the range over which the samples vary that affects accuracy (excluding the possibility that there is an as yet unknown variable responsible for the affect).

In considering whether the affect is due to sample number or to range there are four possible experimental outcomes. One, that range effects accuracy independently of sample-set size and sample-set size has no consistent effect on accuracy. Two, sample-set size has an effect on accuracy independent of range and range has no consistent effect on accuracy. However, given the results of Hinson and Lockhead (1986) and Lockhead and Hinson (1986) this second outcome would seem unlikely. Three, that sample-set size has an effect on accuracy independent of range and that range has an effect on accuracy independent of sample-set size. Four, that sample-set size and range must co-vary in order to have an effect on accuracy. As for two, this would seem unlikely given that Hinson and Lockhead (1986) and Lockhead and Hinson (1986) varied range without varying sample-set size.

Given these four possibilities it would be necessary to do two experiments in order to determine how range and sample-set size effect accuracy. It would be necessary to vary range but not sample-set size, termed Experiment A here, as did Hinson and Lockhead (1986) and Lockhead and Hinson (1986). It would also be necessary to vary sample-set size without varying range, termed Experiment B here. Speculatively, if you were to get an effect for Experiment A and not for Experiment B it would suggest that the decrease observed here, for Experiments 5 and 6 were a result of range not sample-set size. If there was an effect for Experiment B and not for Experiment A then it would suggest that the decrease was a result of sample-set size, not range. If you were to get an effect for both Experiment A and Experiment B then it would suggest that both range and sample set-size effect accuracy and it would not be clear which caused the effect seen in Experiments 5 and 6. However, if there was no effect for either Experiment A or B then it may suggest that sample-set size and range must co-vary in order to have an effect on accuracy. This might suggest that the result seen in Experiments 5 and 6 occurred because here, both range and sample-set size varied.

Here, it was clearly shown that changing the probability of ITC had no impact on matching accuracy independent of changes in sample-set size. However, it is possible that if the sample-set size had been varied and the probability of ITC held constant, that changing the sample-set size may have had no effect on accuracy. If this occurred it would suggest that the effect observed here, in Experiments 3 5 and 6 was a result of covariance in sample-set size and the probability of ITC. Given this possibility the probability of ITC should be kept constant for Experiments A and B, above. This could be done by arranging the order in which stimuli appeared on trial, in a manner similar to that used in Experiments 7 and 10, to vary the probability of ITC. If, for both Experiment A and B, the probability of ITC was held constant and neither sample-set size or range were found to effect accuracy then as mentioned above it is possible that range and sample-set size must co-vary. Alternatively it is possible that it is not the case but that either range and ITC or sample-set size and ITC must co-vary in order for accuracy to be effected.

In conclusion it is clear that here, the number of sample stimuli had an effect on remembering and that this effect was not caused by the covariance of sample number and ITC. However, whether this effect is a product of the number of sample stimuli or of other factors, such as range, which co-vary with sample number is still to be determined.

References

Adamson, C. (1995). *Effects of single and multiple sample stimuli on delayed-matching-to-sample in humans.* Unpublished Master's Thesis, Waikato University, New Zealand.

Bardenhagen, F.J., & Bowden, S.C. (1995). Does knowledge of the response rule determine human performance on an animal test of amnesia. *Neuropsychology, 4,* 573-579.

Bowden, S.C., Benedikt, R., & Ritter, A.J. (1992). Delayed matching to sample and concurrent learning in nonamnesic humans with alcohol dependence. *Neuropsychologia, 30,* 247-435.

Buskist, W.F., & Miller, H.L. (1986). Interaction between rules and contingencies in the control of human fixed-interval performance. *The Psychological Record, 36,* 109-116.

Davison, M, C., & Tustin, R.D. (1978). The relation between the matching law and signal detection theory. *Journal of the Experimental Analysis of Behaviour, 29,* 331-336.

Edhouse, W.V., & White, K.G. (1988a). Sources of proactive interference in animal memory. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 56-70.

Edhouse, W.V., & White, K.G. (1988b). Cumulative proactive interference in animal memory. *Animal learning and Behavior, 16,* 461-46

Ellis, H. C. (1978). *Fundamentals of human learning, memory and cognition.* Iowa: W.C. Brown Company Publishers.

Etkin, M., & D'Amato, M.R. (1969). Delayed matching-to-sample and short term memory in the Capuchin monkey. *Journal of Comparative and Physiological Psychology, 69*, 544-549.

Ferguson, G.A. (1965). *Nonparametric trend analysis.* Montreal: McGill University Press.

Foranzo, L.B., & Logue, L.W. (1995). Self-control and impulsiveness in children and adults: Effects of food preferences. *Journal of the Experimental Analysis of Behavior, 64*, 33-46.

Galizio, M., & Buskist, W. (1988). Laboratory lore and research practices in the experimental analysis of human behavior: selecting reinforcers and arranging contingencies. *The Behavior Analyst, 11*, 65-69.

Grant, D.S. (1975). Proactive interference in pigeon short-term memory. *Journal of Experimental Psychology, 104*, 207-220

Gravetter, F., & Lockhead, G.R. (1973). Criterial range as a frame of reference for stimulus judgement. *Psychological Review, 80*, 203-216.

Hart, R.P., Kwentus, J.A., Taylor, J.R., & Harkins, S.W. (1987). Rate of forgetting in dementia and depression. *Journal of Consulting and Clinical Physchology,* **5,** 101-105.

Hautus, M.J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments & Computers, 27,* 46-51.

Herrnstein, R.J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4*, 267-272.

Hinson, J.M., & Lockhead, G.R. (1986). Range effects in successive discrimination. *Journal of Experimental Psychology, 12*, 270-276.

Holdstock, J.S., Shaw, C. & Aggleton, J.P. (1995). The performance of amnesic subjects on tests of delayed matching-to-sample and delayed matching-to-position. *Neuropsychologia, 33*, 1583-1596.

Huppert, F.A., & Piercy, M. (1978). Dissociation between learning and remembering in organic amnesia. *Nature, 275*, 317-318.

Huynh. H., & Feldt, L.S. (1970). Conditions under which mean square ratios in repeated measures designs have exact *F*-distributions. *Journal of the American Statistical Association, 65,* 1582-1589.

Kopelman, M.D. (1985). Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome. *Neuropsychologia, 23*, 623-638.

Kopelman, M.D. (1992). Storage, forgetting and retrieval in the anterograde and retrograde amnesia of Alzheimer's disease. In L. Backman (Ed.), *Memory functioning in dementia.* (pp. 45-73).

Lockhead, G.R., & Hinson, J. (1986). Range and sequence effects in judgement. *Perception and Psychophyics, 40*, 53-61.

Martone, M., Butters, N. and Trauner, D. (1986). Some analyses of forgetting of pictorial material in amnesic and demented patients. *Journal of Clinical and Experimental Neuropsychology, 8*, 161-178.

Mason, M., & Wilson, M. (1974). Temporal differentiation and recognition memory for visual stimuli in Rhesus monkeys. *Journal of Experimental Psychology, 103,* 383-390.

McCarthy, D. (1991). Behavioral Detection Theory: some implications for applied human research. In M.L. Commons, J.A. Nevin, & M.C. Davison. (Eds.), Quantitative analysis of behaviour: Vol 11. *Signal Detection: mechanisms, models and applications.* (pp. 239-253). Hillsdale, NJ: Erlbaum Associates.

McCarthy, D., & White, K.G. (1987). Behavioral models of delayed detection and their application to the study of memory. In M.L. Commons, J.E. Mazure, J.A. Nevin, & H.Rachlin (Eds.), Quantitative analysis of behaviour: Vol. 5. *The effect of delay and of intervening events of reinforcement value* (pp 29-51.) Hillsdale, NJ: Lawrence Erlbaum Associates.

McCarthy, D., & Davison, M. (1980). Independence of Sensitivity to relative reinforcement rate and discriminability in signal detection. *Journal of the Experimental Analysis of Behavior, 34*, 273-284.

Mishkin, M., Delacour, J. (1975). An analysis of short-term visual memory in the monkey. *Journal of Experimental Psychology: Animal Behavior processes, 1*, 326-334.

Moise, S.L. (1976). Proactive effects of stimuli, delay and response position during delayed matching to sample. *Animal Learning and Behavior, 4*, 37-40.

Money, E.A., Kirk, R.C., & McNaughton, N. (1992). Alzheimer's dementia produces a loss of discrimination but no increase in rate of memory decay in delayed matching to sample. *Neuropsychologia, 30,* 133-43.

Morris, R.G. (1992). Patterns of short-term memory impairment in Alzheimer's disease. In Backman, L. (Ed.), *Memory functioning in Dementia* (pp. 3-25). Amsterdam: North-Holland.

Nevin, J.A., & Grosch, J. (1990). Effects of signalled reinforcer magnitude on delayed matching-to-sample performance. *Journal of Experimental Psychology, 16*, 298-305.

Newman, B., Buffington, D.M., & Hemmes, N.S. (1995). The effects of schedules of reinforcement on instruction following. *The Psychological Record, 45*, 463-476.

Parr, W.V. (1992). Delayed matching-to-sample performance as a measure of human visuospatial working memory. *Bulletin of the Psychonomic Society, 30*, 369-372.

Peterson, L.R., & Peterson, M.J. (1959). Short-tem retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-198.

Roberts, W.A. (1980). Distribution of trials and intertrial retention in delayed matching to sample with pigeons. *Journal of Experimental Psychology, 6,* 217-237.

Ruske, A.C., Fisher, A., & White, K.G. (1997). Attenuation of scopolamine-induced deficits in delayed-matching performance by a new muscarinic agonist. *Psychobiology, 24*, 313-320.

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

Spiga, R., Cherek, D.R., Grabowski, J., & Bennet, R.H. (1992). Effects of inequity on human free-operant cooperative responding: a validation study. *The Psychological Record, 42*, 29-40.

Stromer, R., McIlvane, W.J., Dube, W.V., & MacKay, H.A. (1993). Assessing control by elements of complex stimuli in delayed matching to sample. *Journal of the Experimental Analysis of Behavior, 59*, 83-102.

Tarr, M.J., Kersten, D., & Bülthoff, H.H. (1998). Why the visual recognition system might encode the effects of illumination. *Vision Research, 38,* 2259-2275.

Trigo, E., Martinez, R., & Moreno, R. (1995). Rule performance and generalization in a matching-to-sample task. *The Psychological Record, 45,* 223-240.

Ward, L.M., & Lockhead, G.R. (1971). Response system processes in absolute judgement. *Perception and Psychophysics, 9,* 73-78.

Weardon, J. H. (1988). Some neglected problems in the analysis of human operant behavior. In G. Davey, & C. Cullen (Eds.), *Human operant conditioning and behavior modification* (pp. 197-224). Chichester: John Wiley and Sons.

Wernham, C. J. (1997). *Automatic control of stimulus disparity as a preparation for delayed matching to sample.* Unpublished Master's Thesis, Waikato University, New Zealand.

White, K.G. (1985). Characteristics of forgetting functions in delayed matching to sample. *Journal of the Experimental Analysis of Behavior, 44*, 15-34.

White, K.G. (1991). Psychophysics of direct remembering. In Commons, M.L., Nevins, J.A. & Davison, M.C (Eds.), *Signal detection mechanisms, models and applications* (pp. 221-237). New Jersey: Lawrence Erlbaum Associates.

White, K.G. (2001). Forgetting Functions. *Animal Learning and Behavior, 29,* 193-207.

White, K.G., & McKenzie, J. (1982). Delayed stimulus control: recall for single and relational stimuli. *Journal of the Experimental Analysis of Behavior, 38,* 305-312.

White, K.G., & Alsop, B. (1993). Cognition in birds. In A. Sahgal (Ed.), The practical approach series: V1. *Neuroscience* (pp. 137-149). Oxford: Oxford University Press.

White, K.G., Ruske, A.C., & Colombo, M. (1996). Memory procedures, performance and processes in pigeons. *Cognitive Brain Research, 3,* 309-317.

White, K.G., & Wixted, J.T. (1999). Psychophysics of remembering. *Journal of the Experimental Analysis of Behavior, 71,* 91-113.

Wilson, K.G., & Hayes, S.C (1996). Resurgence of derived stimulus relations. *Journal of the Experimental Analysis of Behavior, 66,* 267-281.

Wixted, J.T., & Ebbesen, E.B. (1991). On the form of forgetting. *Psychological Science, 6,* 409-415.

Wixted, J.T., & Dougherty, D.H. (1996). Detecting a nonevent: delayed presence-versus-absence discrimination in pigeons. *Journal of the Experimental Analysis of Behavior, 65,* 81-92.

Worsham, R.W. (1975). Temporal discrimination factors in the delayed matching-to-sample task in monkeys. *Animal Learning and Behavior, 3,* 93-97.

Wurster, R.M., & Griffiths, R.R. (1979). Human concurrent performances: variation of reinforcer magnitude and rate of reinforcement. *The Psychological Record, 29*, 342-354.

Appendix A

The derivation of log *d*

Log *d* is a measure of discriminative behaviour which was originally proposed by Davison and Tustin (1978) as an extension of Herrnstein's (1961) matching law. In a typical detection task a subject is presented with one of two discriminative stimuli and is required to select which of the two stimuli was presented (McCarthy & White 1987). Davison and Tustin (1978) suggested that a detection task could be viewed as two concurrent reinforcement-extinction schedules, where the schedule in effect is signalled by the presentation of one of two discriminative stimuli. They applied the generalised matching law to the stimulus-response matrix shown in Figure *X*, where *w* (hit) is a correct response in the presence of S_1, *z* (correct rejection) is a correct response in the presence of S_2, *x* (false rejection) is an incorrect response in the presence of S_1 and *y* (false alarm) is an incorrect response in the presence of S_2. Davison and Tustin outlined two equations which were an extension of the generalised matching law and also provided a measure of discriminability. The first equation which describes behaviour in the presence of S_1 is:

$$\log\left(\frac{P_w}{P_x}\right) = a_{r1}\log\left(\frac{R_w}{R_z}\right) + \log c + \log d \tag{4}$$

The second equation which describes behaviour in the presence of S_2 is:

$$\log\left(\frac{P_y}{P_z}\right) = a_{r2}\log\left(\frac{R_w}{R_z}\right) + \log c - \log d \tag{5}$$

Where P is the number of responses emitted, R is the number of reinforcers obtained and the subscripts (*w, x, y, z*) refer to the cells shown in Figure *X*. The parameters a_{r1} and a_{r2} measure sensitivity of behaviour allocation in response to changes in the reinforcement ratio for each of the two discriminative stimuli. Log *c* is a measure of inherent bias, or a bias which remains constant regardless of changes in reinforcement or stimuli. Log *d* is a measure of the extent to which the two stimuli are discriminable. The log *d* value will increase as a subject becomes better able to discriminate between the two stimuli. Log *d* has a positive sign in

equation one and a negative sign in equation two because the numerator in both equations is an S_1 response. McCarthy and Davison (1980) showed that a_{r1} and a_{r2} are equal and thus that it was possible to subtract Equation (*X*) from Equation (*X*) to give a bias-free measure of stimulus discriminability, Log *d*:

$$\log d = 0.5\left(\log\left(\frac{P_w P_z}{P_x P_y}\right)\right) \tag{6}$$

Equations (*X*) can be added to Equation (*X*) to give the following equation:

$$\log\left(\frac{P_w}{P_x}\right)+\log\left(\frac{P_y}{P_z}\right)=2a_r\log\left(\frac{R_w}{R_z}\right)+2\log c \tag{7}$$

This equation can be rearranged to give:

$$0.5\log\left(\frac{P_w P_y}{P_x P_z}\right)= a_r\log\left(\frac{R_w}{R_z}\right)+\log c \tag{8}$$

The left-hand side of the equation is a measure of total response bias which is said to be independent of stimulus discriminability and was referred to by McCarthy (1991) as log *b*. The right-hand side of the equation is an expression of inherent bias, log *c*, and reinforcement bias of the bias which occurs in responding due to changes in the ratio of obtained reinforcement. When the ratio of reinforcement for the two stimuli are equal, that is R_w equals R_z, the ratio of R_w to R_z will be equal to 1.00 and the logarithmic ratio will be equal to zero. The zero can then be substituted for the expression of reinforcement bias, on the right hand side of equation x. Thus, response bias, log *b*, will be equal to log *c*, or inherent bias.

Appendix B

The sequence in which the two sample stimuli appeared, in Experiment 10, for each block of trials for each probability of intertrial correspondence.

Trial number	Sample	Trial Pair	Trial Type
1	S1	not analysed	not analysed
2	S2	1	NCT
3	S2	2	CT
4	S1	3	NCT
5	S1	4	CT
6	S2	5	NCT
7	S2	6	CT
8	S1	7	NCT
9	S1	8	CT
10	S2	9	NCT
11	S2	10	CT
12	S1	11	NCT
13	S1	12	CT
14	S2	13	NCT
15	S2	14	CT
16	S1	15	NCT
17	S1	16	CT