Published in final edited form as:

J Exp Anal Behav. 2013 July ; 100(1): 5–26. doi:10.1002/jeab.32.

MATCHING-TO-SAMPLE PERFORMANCE IS BETTER ANALYZED IN TERMS OF A FOUR-TERM CONTINGENCY THAN IN TERMS OF A THREE-TERM CONTINGENCY

Brent M. Jones¹ and Douglas M. Elliffe²

¹Curtin University, Australia ²The University of Auckland, New Zealand

Abstract

Four pigeons performed a simultaneous matching-to-sample (MTS) task involving two samples and two comparisons that differed in their pixel density and luminance. After a long history of reinforcers for correct responses after both samples, 15 conditions arranged either continuous reinforcement of correct responses after Sample 1 and extinction for all responses after Sample 2, or vice versa. The sample after which correct responses were reinforced alternated across successive conditions. The disparity between the samples and the disparity between the comparisons were varied independently across conditions in a quasifactorial design. Contrary to predictions of extant quantitative models, which assume that MTS tasks involve two 3-term contingencies of reinforcement, matching accuracies were not at chance levels in these conditions, comparison–selection ratios differed after the two samples, and effects on matching accuracies of both sample disparity and comparison disparity were observed. These results were, however, consistent with ordinal and sometimes quantitative predictions of Jones' (2003) theory of stimulus and reinforcement effects in MTS tasks. This theory asserts that MTS tasks involve four-term contingencies of reinforcement and that any tendency to select one comparison more often than the other over a set of trials reflects meaningful differences between comparison-discrimination accuracies after the two samples.

Keywords

matching-to-sample; signal detection; discriminated operant; quantitative models; reinforcement contingencies; key peck; pigeons

The publication of signal-detection theory (Green & Swets, 1966) not only influenced the research conducted by psychophysicists, it also inspired behavior analysts to develop quantitative models of behavior in tasks where stimulus discriminations are taught (Nevin, 1969). Researchers in both camps recognized that the accuracy with which a participant performs a detection (or discrimination) task could be affected by *motivational* variables that determine bias for either response in the task, and *stimulus* variables that determine how able the subject is to detect a stimulus, or sense a difference between stimuli. Signal-detection theory offered independent measures of response bias and stimulus detectability with which to identify those variables. However, some behavior analysts developed new quantitative models that offered alternative measures. Known collectively as *behavioral-detection*

Address correspondence to Max Jones, School of Psychology & Speech Pathology, Curtin University, GPO Box U1987, Perth, WA 6845, Australia (brent.jones@curtin.edu.au).

Please communicate about this manuscript with: Max Jones, School of Psychology & Speech Pathology, Curtin University, Perth, Western Australia, AUSTRALIA, Ph: (+618) 9266 2468, brent.jones@curtin.edu.au

theories (BDTs), these models adopted quantitative models of choice on concurrent schedules of reinforcement (e.g., Baum's 1974 generalized matching law; Davison & Jenkins' 1985 contingency-discriminability model) as theories of response bias in detection and discrimination tasks.

Researchers assessing BDTs have studied steady-state performance in a range of discrete-trial discrimination tasks for nonhumans. These tasks generally fall into one of two categories: analogues of the human "yes"–"no" signal-detection (SD) task and matching-to-sample (MTS) tasks. In the SD paradigm, a so-called *sample stimulus* appears alone on a central display (e.g., a central pecking key for pigeons) at the start of a trial, and a response to it lights up the two side-displays for the choice phase. Operation of one manipulandum (e.g., the left of two yellow keys) is reinforced when one sample appears on the central display, and operation of the other manipulandum (e.g., pecking the right yellow key) is reinforced when the other sample appears. In the MTS paradigm, two *comparison stimuli* are presented simultaneously on the side keys after presentation of a sample; thus, the two side-keys now present different stimuli (e.g., red and green fields). On Sample 1 (S_1) trials, a response to whichever side-key presents Comparison 1 (C_1) is reinforced, and on Sample 2 (S_2) trials, a response to whichever side-key presents Comparison 2 (C_2) is reinforced.

Despite MTS tasks involving an additional set of stimuli (i.e., comparisons), authors of BDTs (e.g., Alsop, 1991; Davison, 1991; Davison & Jenkins, 1985; Davison & Tustin, 1978; Nevin, Davison & Shahan, 2005; Nevin, Jenkins, Whittaker & Yarensky, 1982) have pursued a single quantitative model of performance in SD and MTS paradigms because they have assumed that both paradigms involve conditional discriminations and three-term contingencies of reinforcement. [For example, Nevin et al. (2005) wrote: "The conditional-discrimination paradigm encompasses both matching-to-sample and signal detection, and our model applies to both tasks." (p. 284).] Specifically, the sample stimuli are assumed to function as discriminative stimuli in both paradigms; they signal the availability of reinforcers for operating specific manipulanda in SD tasks, and they signal reinforcers for selecting specific comparisons in MTS tasks. (See Table 1 in Davison & Nevin, 1999). Thus, responses have been defined in terms of their spatial locations in SD tasks, but in terms of the comparison stimulus selected (irrespective of spatial location) in MTS tasks.

This definition of responses in MTS, and the assertion of three-term contingencies of reinforcement in these tasks, have never been questioned by authors of BDTs and perhaps because similar effects of manipulating reinforcement rates were reported in SD and MTS tasks. Specifically, just as varying the relative rate of reinforcement for correct-left responses in a SD task induced varying degrees of left-response bias (e.g., McCarthy & Davison, 1979; Nevin et al. 1982), so too varying the relative rate of reinforcement for correct C_1 responses in a MTS task produced varying degrees of *bias* for selecting C_1 (e.g., Godfrey & Davison, 1998; Harnett, McCarthy & Davison, 1984; Jones & White, 1992). This latter finding suggested that subjects were indeed choosing a comparison on each trial and that that choice was controlled in part by the recent history of reinforcement associated with each comparison. However, assuming three-term contingencies in a quantitative model of MTS raises several theoretical issues and conflicts with views advanced by researchers studying other aspects of MTS performance. With regard to issues, treating a difference between C_1 and C_2 selection frequencies as response bias renders ambiguous the interpretation of any tendency to emit left responses more often than right responses (or vice versa) in MTS tasks where trial scheduling ensures that the two positions are correct equally often. *Position biases* have varied systematically with manipulations of reinforcement variables (e.g., Alsop & Jones, 2008; Jones, 2003; Jones & White, 1992; Katz, 1989; McCarthy & Davison, 1991) and often appear early in MTS training (Cumming & Berryman, 1965; Jackson & Pegram, 1970; Kangas & Branch, 2008). Some researchers

have considered them a second type of response bias (Brown & White, 2009; Katz, 1989; Nevin & Grosch, 1990) but BDTs do not accommodate position bias despite it having implications for bias-free measures of stimulus control (Brown & White, 2009). In addition, the definition of response bias adopted in BDTs discounts the possibility that matching accuracies might on occasions differ on S_1 and S_2 trials. (More accurate matching on S_1 than S_2 trials will result in C_1 having been selected more often than C_2). A number of authors have found it useful to postulate different matching accuracies after the two samples in a MTS task (e.g., Cumming & Berryman, 1965; Grant, 1991; Jones & Davison, 1998; Maki, Gillund, Hague & Siders, 1977; Spetch & Wilkie, 1982; Wixted, 1993; Wixted & Gaitan, 2004), but all BDTs view these differences as spurious and reflecting only degrees of a comparison-selection bias. Thus, it is unclear whether a comparison-selection bias produces an accuracy differential or vice versa.

With regard to other views of the reinforcement contingencies in MTS, several researchers (e.g., Cumming & Berryman, 1965; Jones, 2003; Sidman, 1986; 2000) have proposed that MTS tasks can involve four-term contingencies. In these conceptualizations, samples serve a function that is distinct from that served by the comparisons; namely, samples do not control responses directly but, instead, determine the control over responding that is exerted by the comparison stimuli. Thus, samples have been called *conditional stimuli* and identified as the fourth term in the contingencies because they signal which three-term contingency - involving the comparisons as *discriminative stimuli* - is operating on a trial. Moreover, Jones offered a particular four-term conceptualization that was supplemented with a theory to account for the effects of relative rates of reinforcement in MTS tasks described above. This paper reports the results of an experiment that pitted some predictions of Jones' four-term contingency view against those of the three-term contingency view adopted in BDTs.

Jones (2003) adopted Cumming and Berryman's (1965) and Sidman's (1986, 2000) definitions of *simple* and *conditional discriminations* where the former involve invariant three-term contingencies of reinforcement while the latter involve four-term contingencies and, thus, varying three-term contingencies. Jones argued that responses could (and should) be defined the same way in SD and MTS tasks, and that doing so implies that SD tasks involve only simple discriminations between the samples (albeit *two* 3-term contingencies) whereas MTS tasks can involve conditional discriminations between the comparisons. His conceptualization of the reinforcement contingencies in SD and MTS tasks was as follows. In both paradigms, responses were defined with respect to which manipulandum was operated (e.g., left-key vs. right-key pecks). However, the functions afforded to stimuli differed in the two paradigms. Specifically, samples in a SD task were seen as the discriminative stimuli for left and right responses, whereas in MTS, the two comparison configurations (e.g., C_1 on the left and C_2 on the right versus C_2 on the left and C_1 on the right) were offered as the discriminative stimuli for these responses, and samples were conditional stimuli that determined the discriminative function of comparison configurations.

In addition to offering new generic definitions of the stimuli and responses in MTS, Jones (2003) proposed that the samples in MTS serve a second function whenever the two correct responses are reinforced at different rates; that is, as well as signaling which three-term contingency is operating, samples also signal the recent rate of reinforcement for responding in accordance with that contingency and, consequently, acquire stimulus control over observing (or attending to) the comparison array. He assumed that observing the comparisons occurs in an all-or-none manner, and that the probability of comparison observation varies with the rate of reinforcement obtained for correct responses that follow observing. Different rates of reinforcement on S_1 and S_2 trials will thereby engender different probabilities of observing the comparisons on these trials, and different

probabilities of observing the comparisons will, in turn, give rise to different comparison-discrimination accuracies (i.e., the degree to which responding on the two comparison configurations differs). Thus, in contrast to BDTs, Jones proposed that different comparison-discrimination accuracies after the two samples cause a difference in the overall frequency with which the two comparisons are selected, and so mimic a comparison-selection bias. A secondary aim of the present experiment was to seek evidence for different probabilities of observing the comparisons after the two samples in analyses of dependent measures other than comparison-discrimination accuracy; namely, analyses of latencies to peck at a comparison.

In this experiment, pigeons performed a simultaneous MTS task involving two samples and two comparisons. In all conditions, continuous reinforcement (CRF) was arranged for correct responses after one sample and extinction (Ext) was arranged after the other sample, but the sample (and comparison) that was associated with CRF alternated across successive conditions. The physical disparity between the samples, and the disparity between the comparisons, were then varied across conditions in a quasifactorial design. BDTs predict specific effects of these manipulations. These predictions follow from their three-term contingency view and can be shown when their equations are solved with reinforcers obtained after either S_1 or S_2 set to zero. All BDTs predict that subjects will select the reinforced comparison to an equal extent on S_1 and S_2 trials in all conditions. This will result in matching accuracies being at chance levels (i.e., proportion-correct scores $\approx .5$) throughout. Moreover, they all predict no effect of sample disparity on measures of bias for selecting the reinforced comparison. Predictions of the various BDTs differ with respect to only the degree of this bias and whether it should be affected by comparison disparity. According to Davison and Tustin's (1978) model, subjects should select the reinforced comparison exclusively in all conditions, independent of both disparities.¹ Consequently, their measure of response bias should be either zero or infinite because it utilizes ratios of response types. Alsop's (1991), Davison's (1991) and Nevin et al.'s (2005) models, on the other hand, accept that subjects might occasionally select the comparison associated with extinction when the disparity between comparisons renders the responses less than perfectly discriminable. Thus, these latter models predict that comparison-selection biases will sometimes be less than exclusive and should fall with decreasing comparison disparity.²

In contrast to BDT predictions, Jones' (2003) theory predicts that overall matching accuracies will be above chance levels whenever sample and comparison disparities exceed a perceptual threshold, and that there will be main effects of both disparities on overall accuracy. His theory asserts that the three-term contingency in force on a trial, and the recent rate of reinforcement for responding in accordance with that contingency, will both be signaled clearly by the samples in conditions where sample disparity is high. Consequently, observing will occur on almost every trial involving CRF and almost none of the trials involving Ext. Therefore, comparison-discrimination accuracy will be high and determined by the disparity between the comparisons on CRF trials, but be at chance levels and independent of comparison disparity on Ext trials. Given that CRF (e.g., S_1 trials) and Ext trials (e.g., S_2 trials) will occur equally often, averaging comparison-discrimination accuracies on the two trial types in conditions arranging high sample and comparison

¹This prediction results from Davison and Tustin's use of Baum's (1974) generalized matching law as their underlying model of choice. Emission of only one of the available responses is predicted by this model of choice when the obtained reinforcer ratio over a defined period is either $R_1/0$ or $0/R_2$ irrespective of the values of the free parameters a and c .

²These models predict an effect of comparison disparity on their measure of response bias because they each apply Davison and Jenkins' (1985) model of choice and that model incorporates a free-parameter to measure the discriminability between responses (d_r , Alsop, 1991; Davison, 1991) or the discriminability between response-reinforcer relations (d_{br} , Davison & Nevin, 1999; Nevin et al, 2005).

disparities will result in overall matching accuracies being around .75 correct (i.e., near 1.0 correct on CRF trials and around .5 correct on Ext trials).

Jones' (2003) theory also predicts that overall matching accuracies will fall with decreasing sample disparity because the two signaling functions of the samples will have been degraded by this manipulation. This decrease in overall accuracy should, however, be manifest as a decrease in comparison-discrimination accuracy on only the Ext trials; accuracies on CRF trials should remain high and constant. To understand why his theory predicts such asymmetrical effects, it is helpful to consider the extreme case where the two samples are identical and so must lose both signaling functions. When correct responses after only one sample are reinforced but subjects cannot discriminate the samples, four-term contingencies of reinforcement must reduce to three-term contingencies. That is, the procedure should now be analogous to arranging a single SD task (i.e., two 3-term contingencies of reinforcement) albeit with a sample-key response required to initiate each trial, comparison configurations serving as the discriminative stimuli, and reinforcers available unpredictably on half of the trials. If this statement is correct, and predictions regarding performance at high sample and comparison disparities are also correct, then, subjects should behave on an increasing proportion of the Ext trials in the same manner as they behave on CRF trials as sample disparity decreases. They should observe the comparisons on these *misperceived* Ext trials, but the wrong pair of three-term contingencies will also have been signaled and will result in their emitting the wrong (left or right) response. Some number of incorrectly performed Ext trials will thereby be combined in analyses with the Ext trials on which observing failed to occur, the total number of errors scored on Ext trials will increase, and comparison-discrimination accuracy on aggregated Ext trials will drop below chance level. Moreover, this decrease in discrimination accuracy on only Ext trials will result in subjects selecting the reinforced comparison on an increasing proportion of total trials as sample disparity falls. Thus, Jones' theory predicts that the measure of response bias used by BDTs should increase with decreasing sample disparity whereas BDTs themselves predict no change.

Method

Subjects

Four ex-homing pigeons, numbered 11 to 14, were maintained at $85\% \pm 15$ g of their free-feeding body weights by supplementary feeding of mixed grain after each training session. All pigeons had previously served in experiments reported by Jones (2003) and so had extensive training on a MTS task that involved the same apparatus as that used here.

Apparatus

Four identical cages served as both home cages and the environments in which training sessions were conducted. These cages were situated in a large room containing around 90 other similar cages serving unrelated experiments and artificially lit to provide a 16h-8h light-dark cycle. No person entered this room while sessions were in progress.

Each cage measured 370 mm wide, 380 mm deep and 380 mm high. The left and rear walls were constructed of iron sheets, whereas the floor, ceiling, and front wall consisted of iron rods spaced 50 mm apart. Two wooden perches were arranged inside the cage to enable access to water and grit outside the cage, and access to an interface panel mounted on the right wall. The interface panel consisted of three response keys and an aperture through which wheat could be delivered. The response keys were made of transparent plexiglass, were 25 mm in diameter, were arranged horizontally (47 mm between centers) and central on the panel, and were mounted 215 mm above the perch. Pecks on lit keys were registered when they exceeded about 0.1 N. An aperture measuring 52 mm wide and 52 mm high was

located below the center key and 65 mm above the perch. Presentations of the food hopper through this aperture were accompanied by illumination of the aperture by yellow LEDs and the extinction of all key stimuli.

A monochromatic liquid-crystal display (LCD) of the sort used in electronic typewriters was mounted behind the response keys. This LCD presented black pixels (0.49 mm square) on a green background, and was backlit with white light. Three equally sized regions of the display were controlled independently so as to present different images behind each key. However, the backlight illuminated the entire display and, therefore, all three keys simultaneously. About 1452 pixels could be viewed through each response key. The imaging surface of the LCD was recessed 6 mm from the rear of the keys, the effect of which was to reduce the viewing angle of stimuli in each position relative to the angle available when stimuli are projected onto opaque keys (cf. Wright & Sands, 1981).

The LCD, food hopper, and hopper light were controlled by an IBM-compatible computer running a program written in TurboPASCALO™ and situated outside the room housing experimental cages. This computer also recorded the time and type of all stimulus events and pecks to lit keys within a session.

Procedure

Each pigeon received one training session per day, seven days per week. The sessions for each pigeon were run successively and while other unrelated experiments involving other pigeons were running. Sessions were initiated by the controlling computer and occurred at approximately the same time each day.

The procedure used here was identical to that described by Jones (2003) except that we arranged different schedules of reinforcement for correct responses, and lower sample and comparison disparities. In view of their extensive pretraining histories, pigeons were introduced immediately to Condition 1 of the present experiment.

Each session in all conditions involved a simultaneous MTS task involving two samples and two comparisons. A trial started with the illumination of the LCD by the backlight, and the appearance of a random-dot pattern—a sample stimulus—on the center key. Patterns with relatively few activated pixels were designated instances of S_1 and were generated when the computer swept across the central third of the LCD and activated each pixel in that area with the probability given in Table 1. Patterns with relatively many activated pixels were instances of S_2 and generated when each pixel in the central area was activated with a higher probability. The selection of S_1 or S_2 for each trial was random. A single center-key peck in the presence of either sample caused two other random-dot patterns to appear on the side keys. The pattern on one side-key was designated C_1 and was generated by activating pixels with a relatively low probability (see Table 1). The pattern on the other side key was designated C_2 and was generated with a higher probability (see Table 1). The location of the two patterns (C_1 and C_2) behind the side keys was randomized across trials such that each comparison appeared on each side key approximately equally often in a session. A peck on either side key extinguished the backlight and deactivated all pixels on the LCD. The additional consequence of a side-key peck depended upon which comparison had appeared on that key, which sample had appeared on that trial, and the experimental condition in effect. Correct responses were defined as pecks to the side-key displaying C_1 on S_1 trials and pecks to C_2 on S_2 trials, whereas errors were pecks to C_2 on S_1 trials and pecks to C_1 on S_2 trials. All errors resulted in 3-s time-out from the task. However, only one of the two correct-response types (e.g., pecks to C_1 on S_1 trials) was reinforced in any one condition; the other correct-response type (e.g., pecks to C_2 on S_2 trials) earned the same consequence as errors. Thus, in each condition, continuous reinforcement (CRF) was scheduled for

correct responses on either S_1 trials or S_2 trials, and the three remaining response types (two types of errors and one correct-response type) were extinguished. Reinforcement of the selected response type involved providing 3-s access to the food hopper. Following either food access or time-out, a 5-s intertrial interval ensued before the next trial. Sessions ended after 45 min had elapsed, or after 50 reinforcers had been obtained, whichever event occurred sooner.

Table 1 shows the probabilities that were used to generate the four stimuli, which comparison stimulus was associated with reinforcement (C_1 or C_2), and the number of training sessions that each subject received in each of 15 conditions. Also shown are indices of the sample and comparison disparity arranged in each condition to aid an overview of how each disparity varied across conditions. (Sample-disparity indices are the differences between the probabilities used to generate the two samples and comparison-disparity indices are the differences between the probabilities used to generate the two comparisons.) Four levels of sample disparity and comparison disparity were arranged, and combined in a quasifactorial design. Each condition operated until all subjects had received at least 40 sessions and there were no clear trends in either proportion-correct scores or comparison-selection biases (proportion of trials where C_1 was chosen) over the last 20 sessions of that condition when graphs of both measures were inspected visually.

Although sample and comparison disparity was measured in terms of the pixel densities of LCD images, the luminous intensity of the LCD behind a key varied with pixel density because each activated pixel obscured the backlight emanating from that position on the LCD. Thus, stimulus patterns with relatively low pixel densities (e.g., when .3 of the pixels were lit) were also relatively bright response keys. Therefore, it is possible that key brightness, rather than pixel density, was the stimulus dimension that controlled the responding of subjects when matching accuracies were high. We did not, however, measure the luminous intensities associated with different pixel densities and assert that varying the probability of activating pixels was an effective way to manipulate the disparity between stimuli whichever dimension exerted stimulus control.

Results

The raw data from each subject in each condition are shown in Appendix A. These data represent response and reinforcer tallies for the final 20 sessions of a condition. Response frequencies were transformed by adding 0.5 whenever response ratios were calculated because 15 of the 480 frequencies were zeros. This correction is conventional and was recommended by Hautus (1995).

We conducted three sets of analyses of our data. First, we analyzed overall matching accuracies in order to compare our data with the predictions of BDTs and Jones' (2003) theory. Second, we analyzed the data in terms of the dependent measures used in BDTs, and third, we conducted analyses in terms of the dependent measures offered in Jones' theory. All three sets of analyses used the same data, albeit after they were aggregated in different ways. The statistical significance of changes in each dependent measure was assessed using Ferguson's (1966) nonparametric test for monotonic trends on data from individual subjects and with an alpha level of .05.

Overall Matching Accuracies

Figure 1 shows proportion-correct scores for each subject in each condition plotted as a function of the disparity between the samples at each level of disparity between the comparisons. The effect of sample disparity on overall accuracy can be seen within panels whereas the effect of comparison disparity can be seen by comparing data across panels in a

row. Open circles depict data from conditions that were systematic replications of other conditions. These replication conditions differ from one other condition only in that reinforcers were associated with the other comparison stimulus.

Several results are apparent in Figure 1. First, overall matching accuracies in replication conditions were not systematically different from overall accuracies in their counterpart conditions. (Note that unfilled circles occasionally obscure filled circles plotted behind them.) This finding confirms that overall matching accuracy was independent of which comparison stimulus was associated with reinforcement in a condition, and that steady-state performances were obtained from the final 20 sessions of each condition. Second, contrary to the predictions of BDTs but consistent with those of Jones' (2003) theory, matching accuracies exceeded chance levels (i.e., proportion correct > .5) in most conditions. Thus, comparison choice in most conditions continued to be controlled by the sample presented on a trial despite our having reinforced correct responses following only one sample. Third, Figure 1 shows that proportion correct decreased as both sample and comparison disparities decreased. These trends were confirmed significant by trend tests (see Analyses 1 to 6 in Table 2). Finally, proportion correct averaged .748 across subjects and conditions when sample and comparison disparity was highest (Conditions 1 and 2). This value is very close to that predicted by Jones' theory and, as we show later, it represents near perfect comparison-discrimination accuracy following the sample associated with CRF and only chance-level accuracy following the sample associated with Ext (see Figure 3).

BDT—Three-term Contingency—Analyses

Figure 2 shows how variation of sample and comparison disparity affected the measures of discrimination accuracy ($\log d$) and response bias ($\log b$) offered in Davison and Tustin's (1978) model. These performance measures are often used when other BDT models are being investigated (Nevin et al, 2005) because they can be calculated from a single condition (i.e., a single obtained reinforcer ratio), they offer unbounded measures that are expressed in similar units, and the former measure, $\log d$, is closely related to d' from signal-detection theory (Green & Swets, 1966). $\log d$ is the logarithm (base 10) of the geometric mean of the ratios of correct to incorrect responses on S_1 and S_2 trials, whereas $\log b$ is the logarithm (base 10) of the geometric mean of the ratios of C_1 to C_2 selections on S_1 and S_2 trials. That is,

$$\log d = .5 * \log \left(\frac{B_{C_1|S_1} \cdot B_{C_2|S_2}}{B_{C_2|S_1} \cdot B_{C_1|S_2}} \right) \quad (1)$$

and

$$\log b = .5 * \log \left(\frac{B_{C_1|S_1} \cdot B_{C_1|S_2}}{B_{C_2|S_1} \cdot B_{C_2|S_2}} \right), \quad (2)$$

where $B_{C_1|S_1}$ refers to the number of C_1 selections on S_1 trials, $B_{C_2|S_2}$ to the number of C_2 selections on S_2 trials, and so on.

Figure 2 plots the average values of $\log d$ and $\log b$ across subjects as a function of sample disparity for each comparison disparity (top two rows) and as a function of comparison disparity for each sample disparity (bottom two rows). These averages were calculated using data from all conditions per subject. Absolute values of $\log b$ have been plotted because $\log b$ values were always positive in conditions where responses on S_1 trials were reinforced,

and were always negative in conditions where responses on S_2 trials were reinforced. Thus, these absolute values of $\log b$ measure the degree of bias for selecting whichever comparison was associated with reinforcement in a condition.

Recall that Davison and Tustin's (1978) model predicts that $\log d$ (Equation 1) and $\log b$ (Equation 2) values should be infinite or zero in all conditions. In contrast, Alsop's (1991), Davison's (1991) and Nevin et al.'s (2005) models predict that $\log d$ should not be systematically different from zero in all conditions, and that $\log b$ should indicate a bias toward the comparison associated with CRF (the reinforced comparison), should remain constant with variations of sample disparity, and should fall with a decrease in comparison disparity. Figure 2 provides support for some of these predictions but not others. First, contrary to the predictions of all BDTs, $\log d$ values exceeded zero in all conditions. They were highest when sample and comparison disparity were highest (mean $\log d = 1.09$), and decreased as both sample disparity and comparison disparity decreased. Most of these trends were statistically significant; $\log d$ decreased significantly with decreasing sample disparity when comparison disparity was .4 and .1, but not when it was .2 (see Analyses 7, 8 and 9 in Table 2). In addition, $\log d$ decreased significantly with decreasing comparison disparity at each of the three higher sample disparities (see Analyses 10, 11 and 12 in Table 2).

Second, Figure 2 shows that $\log b$ values varied with *both* sample and comparison disparity. The effect of comparison disparity was as Alsop's (1991), Davison's (1991) and Nevin et al.'s (2005) models predicted; $\log b$ values fell systematically with decreasing comparison disparity at all three sample disparities (see Analyses 16, 17 and 18 in Table 2). Thus, bias for selecting the comparison associated with reinforcement decreased as comparison disparity decreased. However, none of the BDTs predicted that $\log b$ values would often increase with decreasing sample disparity. An increasing trend was significant at the highest and lowest comparison disparities, but not when the comparison disparity was .2 (see Analyses 13, 14 and 15 in Table 2). This change in $\log b$ values was, however, predicted by Jones' (2003) theory.

Jones (2003)—Four-term Contingency—Analyses

Figure 3 presents the results of an analysis that follows from Jones' (2003) conceptualization of the reinforcement contingencies in MTS. An implication of his generic definitions of stimuli and responses in these tasks is that measures of comparison-discrimination accuracy can be calculated for S_1 and S_2 trials separately. Furthermore, in so far as Davison and Tustin's (1978) $\log d$ is a useful measure of discrimination accuracy in tasks involving three-term contingencies (e.g., SD tasks), Jones argued that it can provide the two measures of comparison-discrimination accuracy in a MTS task. Each measure is obtained by calculating the degree to which left/right response ratios differ on the two comparison configurations after one sample, and henceforth they will be called $\log D$ to distinguish them from the conventional calculation of $\log d$. Thus, assuming that Comparison-Configuration 1 ($Conf_1$) involves C_1 on the left key and C_2 on the right key, and Comparison-Configuration 2 ($Conf_2$) involves C_2 on the left key and C_1 on the right key, a measure of comparison-discrimination accuracy ($\log D$) on S_1 trials is given by

$$\log D(S_1) = .5 \cdot \log \left(\frac{B_{Left|S_1 \& Conf_1} \cdot B_{Right|S_1 \& Conf_2}}{B_{Right|S_1 \& Conf_1} \cdot B_{Left|S_1 \& Conf_2}} \right), \quad (3)$$

and comparison-discrimination accuracy on S_2 trials is given by,

$$\log D(S_2) = .5 \cdot \log \left(\frac{B_{\text{Right}|S_2 \& Conf_1} \cdot B_{\text{Left}|S_2 \& Conf_2}}{B_{\text{Left}|S_2 \& Conf_1} \cdot B_{\text{Right}|S_2 \& Conf_2}} \right), \quad (4)$$

where $B_{\text{Left}|S_1 \& Conf_1}$ refers to the number of correct left-key responses given S_1 and $Conf_1$, $B_{\text{Right}|S_1 \& Conf_2}$ refers to the number of correct right-key responses given S_1 and $Conf_2$, etc.

Equation 3 and Equation 4 were used to calculate $\log D(S_1)$ and $\log D(S_2)$ for each subject in each condition. Depending upon which sample signaled reinforcement in a specific condition, these measures were then coded as either comparison-discrimination accuracies on CRF trials [$\log D(\text{CRF})$] and plotted as filled symbols in Figure 3, or comparison-discrimination accuracies on Ext trials [$\log D(\text{Ext})$] and plotted as open symbols. Figure 3 shows these accuracies plotted as a function of sample disparity for each level of comparison disparity. Data from conditions that replicated other conditions are shown as triangles. These replication data were very similar to data from the original conditions.

Figure 3 provides evidence for the main and interactive effects of sample and comparison disparity that were predicted by Jones' (2003) theory. The figure shows that $\log D$ values on CRF trials were high and did not vary systematically as sample disparity decreased, but that they decreased with decreasing comparison disparities. Trend tests confirmed the absence of trends with decreasing sample disparity (see Analyses 19, 20 and 21 in Table 2) and the significance of a decreasing trend with decreasing comparison disparities (see Analyses 22, 23 and 24 in Table 2).

With respect to comparison-discrimination accuracies on Ext trials, Figure 3 shows that $\log D(\text{Ext})$ values were close to zero whenever sample disparity was at its highest (i.e., .4) at all levels of comparison disparity. Thus, varying comparison disparity did not affect comparison-discrimination accuracies on Ext trials when sample disparity was high. However, $\log D(\text{Ext})$ values became progressively more negative as sample disparity was reduced (see Analyses 25, 26 and 27 in Table 2). These negative $\log D(\text{Ext})$ values indicate differences between choice on the two comparison configurations (and so stimulus control by the comparisons), but choice that involves more selections of the incorrect comparison than of the correct one. Moreover, it is worth noting that at the lowest sample disparity and the highest comparison disparity (.06; left column of graphs), $\log D(\text{Ext})$ was similar in value to $\log D(\text{CRF})$ from the same condition, but opposite in sign. This implies that the relative frequency of having selected the reinforced comparison was similar on CRF and Ext trials when sample disparity was at its lowest.

In addition to $\log D(\text{Ext})$ values falling with decreasing sample disparity, Figure 3 shows effects of comparison disparity on $\log D(\text{Ext})$ values at lower sample disparities; namely, $\log D(\text{Ext})$ values approached chance levels [$\log D(\text{Ext}) = 0$] as comparison disparity decreased. That is, $\log D(\text{Ext})$ values increased with decreasing comparison disparity if these accuracies started negative (e.g., when Sample Disparity = .1) and decreased if these values started positive (e.g., Sample Disparity = .4; Pigeon 14). A decrease in absolute values of $\log D(\text{Ext})$ was significant at all three levels of sample disparity (see Analyses 28, 29 and 30 in Table 2) and represents an increasing effect of comparison disparity as sample disparity decreased.

Figure 4 presents an analysis of response bias in terms of Jones' (2003) four-term conceptualization; an analysis of any tendency to peck the left key more often than the right key independent of the comparison positions on a trial. This bias was measured by calculating the log ratio of left- to right-key responses on the two comparison configurations

following either sample (i.e., $\log B_L/B_R$ on S_1 trials and $\log B_L/B_R$ on S_2 trials). These sample-specific position biases were then coded as biases on either CRF trials or Ext trials depending on the condition operating, and plotted in Figure 4 as a function of sample-stimulus disparity for each level of comparison disparity.

Clear differences between position biases on CRF and Ext trials are apparent in Figure 4. First, biases on Ext trials (open circles) were considerably greater than biases on CRF trials (filled circles) in most conditions and only small biases were generally observed on CRF trials. (Absolute values of $\log B_L/B_R$ on Ext trials exceeded those on CRF trials in 40 out of 44 comparisons). Second, biases on CRF trials remained small and did not change systematically as sample disparity decreased (see Analyses 31, 32 and 33 in Table 2). Similarly, biases on Ext trials remained high and did not change systematically as comparison disparity decreased (see Analyses 40, 41 and 42 in Table 2). However, biases on CRF trials increased slightly, but nevertheless systematically, with decreasing comparison disparity at the two highest levels of sample disparity (see Analyses 34, 35 and 36 in Table 2), and biases on Ext trials decreased significantly as sample disparity decreased at each level of comparison disparity (see Analyses 37, 38 and 39 in Table 2). The magnitude of the decrease in biases on Ext trials was such that biases on these trials were nearly as small as those on CRF trials when sample disparity was at its lowest (Sample Disparity = .06). Thus, similar degrees of stimulus control by the comparisons on CRF and Ext trials at low sample disparities (Figure 3) were accompanied by similarly small degrees of position bias on these trial types (Figure 4). More generally, a comparison of Figures 3 and 4 shows that comparison-discrimination accuracy and position bias covaried on Ext trials; biases were usually large when comparison-discrimination accuracies were at chance levels ($\log D(\text{Ext}) \approx 0$), and low when comparison-discrimination accuracies were higher albeit strongly negative because the number of errors greatly exceeded the number of correct responses.

Discussion

In this experiment, pigeons received training in a MTS task where correct responses after only one sample were reinforced in a set of successive sessions (a condition), and both sample disparity and comparison disparity were varied across conditions. This procedure allowed us to compare the accuracy with which two competing conceptualizations of the reinforcement contingencies in MTS tasks predicted pigeons' performances. One conceptualization has been adopted in quantitative models known as BDTs (i.e., Alsop, 1991; Davison, 1991; Davison & Jenkins, 1985; Davison & Tustin, 1978; Nevin et al, 2005; Nevin et al, 1982). According to this view, MTS involves three-term contingencies where the samples serve as discriminative stimuli for responses that are defined in terms of which comparison stimulus was selected. The other conceptualization invokes four-term contingencies of reinforcement and was offered by Jones (2003). This view asserts that the samples serve as conditional stimuli that switch the discriminative function of comparison-stimulus configurations for responses that are defined in terms of which manipulandum was operated. The three-term view predicts that overall matching accuracies in all our conditions should have been at chance levels. In contrast, the four-term view predicts that overall matching accuracies should start at around .75 proportion correct at the highest sample and comparison disparities, and fall with reductions in both sample and comparison disparity. An analysis of proportion-correct scores (Figure 1) supported the four-term predictions and not those of the BDT models. Furthermore, analyses of performance measures used by BDTs (Figure 2) revealed effects of our independent variables that were not predicted by BDTs, whereas analyses of measures afforded by Jones' theory (Figures 3 and 4) revealed effects that were predicted by his theory. We assert that these findings imply serious inadequacies with a three-term conceptualization of the reinforcement contingencies in MTS tasks but provide support for the four-term conceptualization.

We turn now to discussing the results of analyzing performance measures drawn from each conceptualization, starting with how measures adopted by BDTs were affected by our independent variables. Neither the non-zero values of $\log d$ nor the effect of sample disparity on $\log b$ values seen in Figure 2 are easily interpreted by BDTs. $\log d$ should have been zero in all conditions because subjects should have selected the *reinforced* comparison to an equal extent on S_1 and S_2 trials. $\log b$ should have been unaffected by sample disparity because subjects should be no more able to discriminate responses, and so show a greater bias toward that response which is being reinforced, when sample disparity is lower. In fact, our effect of sample disparity on $\log b$ values violates an important assumption underlying all BDTs; that the degree of control exerted by the reinforcer ratio over choice between comparisons (i.e., bias in their models) is independent of the discriminability between the samples. Both effects are, however, predicted by Jones' (2003) theory. First, $\log d$ exceeded zero in most conditions for the same reasons that proportion correct scores exceeded 0.5 in those conditions. Second, $\log b$ increased with decreasing sample disparity because this manipulation caused comparison-discrimination accuracies on only Ext trials to decrease and this, in turn, resulted in an increase in the proportion of all trials involving selections of the *reinforced* comparison.

In addition to correctly predicting changes in measures used by BDTs, Jones' (2003) theory correctly predicted changes in measures derived from his four-term conceptualization; namely, comparison-discrimination accuracies on each trial type [i.e., $\log D(\text{CRF})$ and $\log D(\text{Ext})$ in Figure 3]. That is, $\log D(\text{CRF})$ values were high, fell with decreasing comparison disparity, but were invariant with decreasing sample disparity. In contrast, $\log D(\text{Ext})$ values were at chance levels when sample disparity was high, became more negative as sample disparity decreased, and approached chance levels at lower sample disparities as comparison disparity decreased. These differential effects of sample disparity on CRF- and Ext-trial performances imply that four-term contingencies will indeed reduce to three-term contingencies when correct responses after only one sample are reinforced *and* samples are unable to engender differential responding. Specifically, when Jones' four-term conceptualization is supplemented with the notion that samples can acquire stimulus control over observing the comparisons, the effects of sample and comparison disparities seen in Figure 3 can be explained as follows: First, $\log D(\text{CRF})$ values were high and did not vary with sample disparity because subjects always observed the comparison array on these trials. Observing was maintained at its maximum level by the CRF of correct responses that ensued. However, $\log D(\text{CRF})$ values fell with reductions in comparison disparity because the subjects' ability to respond differentially to the comparison arrays after observing them was determined by comparison disparity. Second, chance-level discrimination accuracies on Ext trials ($\log D(\text{Ext}) \approx 0$) when sample disparity was high resulted from the subject's failure to observe the comparison array on these trials. Observing on Ext trials had extinguished because neither side-key response was reinforced on these trials, and these trials were clearly signaled when sample disparity was at its highest. Third, $\log D(\text{Ext})$ values decreased as sample disparity decreased because subjects behaved on an increasing proportion of Ext trials in the same way that they behaved on CRF trials; the comparison array was observed on these misperceived trials, and the reinforced comparison was chosen. Thus, with decreasing sample disparity, an increasing number of Ext trials involving comparison observation and an incorrect response was added to a decreasing number of Ext trials on which comparison observation did not occur and accuracy was at chance, with the result that the number of errors on Ext trials overall increased and $\log D(\text{Ext})$ decreased.³

³A similar outcome of averaging occurred when comparison-selection frequencies were summed across the two samples and explains why $\log b$ values in Figure 2 increased with decreasing sample disparity; that is, an increasing number of errors on Ext trials was added to a constant number of errors on CRF trials as sample disparity fell, with the effect that the relative number of responses to the *reinforced* comparison also increased.

Predicting the values of $\log D(\text{RFT})$ and $\log D(\text{Ext})$ in a condition where S_1 and S_2 are identical is then straightforward. The data paths in Figure 3 suggest that $\log D(\text{Ext})$ will be the additive inverse of $\log D(\text{CRF})$ because subjects will behave as if a single pair of three-term contingencies—a single SD task—is operating; one comparison configuration (e.g., C_1 -left/ C_2 -right) would signal intermittent reinforcement for one response (e.g., a left response if C_1 is associated with reinforcers) and the other comparison configuration would signal intermittent reinforcement for the other response. These explanations of stimulus-disparity effects on comparison-discrimination accuracies can, therefore, be viewed as the mechanisms by which behavioral control by four-term contingencies of reinforcement reduces to control by three-term contingencies.

The fourth effect shown in Figure 3 and explained by Jones' (2003) theory concerns the interaction between sample and comparison disparities on $\log D(\text{Ext})$ values. Although these values were not affected by comparison disparity at the highest sample disparity, they approached chance levels [i.e., negative values of $\log D(\text{Ext})$ increased] at lower sample disparities when comparison disparity was reduced. Jones' theory asserts that this interaction occurred because reducing comparison disparity when the samples were less than perfectly discriminable rendered subjects less able to respond differentially to the comparison configurations on those Ext trials on which comparison observing occurred, and so less able to respond in accordance with the wrong three-term contingency on those trials. Being less able to respond in accordance with the wrong contingency on a proportion of trials meant that fewer errors would be recorded on those trials and $\log D(\text{Ext})$ values must approach chance levels.

In addition to Figure 3 showing the effects predicted by Jones' (2003) theory, an analysis of response bias in terms of his conceptualization (i.e., Figure 4) revealed orderly effects of sample disparity, and those effects can be interpreted by his theory. In essence, the degree of position bias shown on CRF and Ext trials was determined by the proportion of trials on which comparison observing was predicted to have occurred. That is, the higher the proportion of trials involving observing, the lower the ceiling on measures of position bias calculated when all trials involving one sample type are aggregated. Thus, position biases on all CRF trials were small and were invariant with sample and comparison disparity because subjects were presumed to have observed the comparison array on most (if not all) of these trials in all conditions. In contrast, position biases were largest on Ext trials when sample disparity was at its highest because subjects failed to observe the comparisons on the vast majority of these trials thereby presenting many opportunities for a bias to emerge. Finally, position biases on Ext trials fell with decreasing sample disparity because the proportion of Ext trials on which observing occurred presumably increased as sample disparity decreased.

Given the mediational role of comparison observation in Jones' (2003) theory, a secondary aim of our experiment was to seek supporting evidence for varying proportions of trials on which observing occurred. Our analyses of observing were inspired by analyses conducted by Wright and Sands (1981). These researchers studied observing in a MTS task by constructing a response panel that required pigeons to be directly in front of a key before they could see the stimulus presented there. This enabled human observers to review video records of sessions and judge where a pigeon was *looking* at various times within trials. They summarized analyses of their pigeons' observing as follows: "The pigeon, following its observation of the sample stimulus, orients either to the right or the left and observes that side-key comparison stimulus. It almost always orients in the same direction. If the comparison is judged to match the sample, then it pecks it. If not, it switches to the alternative comparison and repeats the decision process" (p.204). Thus, observing responses were treated as measurable components of a behavior chain that terminated with a peck to one of the comparisons, and analyses of this chain revealed considerable order.

When this characterization of observing is adopted, several features of our results are consistent with comparison-discrimination accuracies (log D values) being determined in part by the proportion of trials on which comparison observing occurred. First, the absence of any effect of comparison disparity on comparison-discrimination accuracies on Ext trials when sample disparity was high (Figure 3) suggests that a failure to observe the comparison array was responsible for chance-level matching accuracies on these trials. That is, if one accepts that a controlling relation between a stimulus and a response can be established by independent stimulus variation and subsequent response measurement (e.g., Honig, 1969; Ray, 1969) then this failure of comparison disparity to affect comparison selection suggests no such controlling relation. Second, our assertion that chance-level accuracies on Ext trials resulted from observing failures is consistent with the presence of strong position biases on these trials. Pigeon 12, for example, pecked the right comparison key on between 89% and 99% of Ext trials, but only 46% to 53% of CRF trials, in conditions arranging the highest sample disparity (see Figure 4). Recall that pigeons in Wright and Sands' (1981) experiments almost always oriented first to the same side key upon presentation of the comparison array. If our pigeons had also developed stereotyped patterns of orienting to side keys, then the position biases on Ext trials imply that their observing on these trials very seldom switched from one position to another before a comparison key was pecked, and that, at best, they observed only one comparison stimulus per Ext trial; namely, that one appearing on the key that they pecked.

An implication of Wright and Sands' (1981) results is that orienting toward one side key on all trials, followed by orienting toward the other side key on 50% of trials, and perhaps orienting back to the first key on some smaller percentage of trials, must occupy some amount of time. This implication inspired us to analyze our subjects' latencies to make a side-key response after presentation of comparisons. We asked whether the distributions of latencies on CRF and Ext trials differed and whether those distributions changed with decreasing sample disparity. Provided that latencies to observe comparisons and peck a side-key on CRF trials are not too short, the absence of observing on Ext trials should result in shorter latencies on Ext trials than on CRF trials. Furthermore, latency distributions on CRF trials should be unaffected by sample disparity and distributions on Ext trials should shift toward those on CRF trials as sample disparity decreased if the proportion of Ext trials that involved observing increased. Although all our subjects showed distribution differences in most conditions and the predicted effects were apparent in mean data, latencies on CRF trials were sufficiently long to avoid a floor effect for only Pigeon 12. Figure 5 shows an analysis of this subject's latencies. The relative frequency with which side-key response latencies fell in bins of 0.2 s was calculated for CRF and Ext trials separately, and these proportions have been plotted for bins up to 4 s, with the right-most bin showing the proportion of responses exceeding 4 s. These relative-frequency distributions have been displayed so that rows of graphs present data from conditions with the same sample disparity but varying comparison disparities, and columns of graphs present conditions with the same comparison disparity but varying sample disparities.

Figure 5 shows that the distributions from CRF trials (drawn with solid lines) resembled normal curves with medians ranging from 1.43 s to 2.03 in all conditions except Condition 14. There were no systematic changes in features of the CRF distributions as either sample disparity or comparison disparity reduced, just as position biases on CRF trials were always low and unaffected by either disparity (Figure 4). In contrast, the distributions of latencies from Ext trials (drawn with dashed lines) in all conditions except Condition 14 were bimodal; one mode was very short (between 0.6 and 0.8 s in 11 of the 14 conditions) and the other exceeded 4 s. Thus, not only did this pigeon show strong position biases on Ext trials in most conditions (Figure 4), but these biased responses occurred at either very short or very long latencies relative to latencies on CRF trials. Long latencies to peck at a

comparison when no primary reinforcer is available on that trial have previously been reported in a MTS procedure (DeMarse & Urcuioli, 1993) but we are not aware of any studies reporting very short latencies under these conditions. Nevertheless, it seems reasonable to assert that these very quick responses represented quick escapes from a stimulus condition that signaled the unavailability of reinforcers. Put another way, the behavior that we describe as observing (i.e., a sequence of orienting to the side keys) had likely not occurred on those Ext trials. In addition, these short Ext-trial latencies were frequently shorter than the shortest latencies seen on CRF trials. It is possible then that this pigeon (and probably others) made these quick position-biased responses without having observed the comparisons behind both keys and perhaps without observing even the comparison on the key that was pecked.

Figure 5 also shows that the shape of the distributions from Ext trials often changed and began to look more like distributions from CRF trials as sample disparity decreased. That is, the relative frequency of very short latencies on Ext trials generally fell, and relatively more responses occurred at intermediate latencies, as sample disparity decreased. Consequently, when sample disparity was at its lowest (Condition 14, lowest panel), the distributions from CRF and Ext trials were nearly indistinguishable. Although this merging of distributions is forced by rendering the samples very difficult to discriminate and so might seem trivial, the fact that only the Ext-trial distributions changed with sample disparity is consistent with our assertion that there was an increase in the proportion of Ext trials on which comparison observing occurred as sample disparity fell. This increase in observing on Ext trials was such that observing occurred on most (if not all) Ext trials at the lowest sample disparity.

Before leaving a discussion of observing, we think it useful to describe how our treatment of observing compares with the treatments offered in previous studies of conditional-discrimination performance and in Nevin et al.'s (2005) BDT. With respect to the former, numerous authors have proposed that some proportion of the errors made in a discrimination task result from failing to observe (or attend to) the relevant stimuli (e.g., Blough, 1996; Gibbon & Church, 1984; Heinemann, Chase & Mandell, 1968). However, most of these researchers have considered lapses in observing only the sample stimulus. For example, Carter and Werner (1978) wrote "...pecking center and side keys per se and matching accurately appear to be under the control of different variables. They do not necessarily occur together, ... (and) ... one cannot equate pecking the center key with attention to the sample" (p.581). We think it useful to supplement their statement with "...and neither can one equate pecking a side key with attention to the comparisons."

Nevin et al. (1982) reported within-session effects of reinforcement rates that are similar to ours. However, they arranged a SD task in which left-key responses were reinforced on trials where a central key had been lit for 3 s and right-key responses were reinforced when the central key had been lit for 2 s. In Experiment 2, they used auditory stimuli—a tone & a noise—to signal different rates of reinforcement for correct responding on a trial. Half of the trials in a session had the tone playing for their duration, and the other half had the noise. The scheduled probability of reinforcement was always five times greater on "tone" trials than on "noise" trials, and the scheduled probabilities on tone trials ranged from 1.0 to .2 across conditions. Nevin et al. found that a measure of accuracy that is algebraically equivalent to $\log d$ (Equation 1) was higher on tone trials than on noise trials in 92% of comparisons, and that across conditions "all subjects exhibited a rough positive relation between ... (this measure)... and the number of reinforcers obtained per trial" (p. 73). Although they did not discuss their results in terms of observing or attention, it seems reasonable to conclude that their pigeons observed the duration of the center-key light on a greater proportion of tone trials than noise trials in all conditions, and that the proportion of trials on which observing occurred was some function of the probability of reinforcement for

correct responding. This is the same mechanism that we propose is operating in MTS tasks when the S_1/S_2 reinforcer ratio is varied.

The critical difference between Nevin et al.'s (1982) procedure and a MTS task is that the auditory stimuli in their procedure served only one function (i.e., they signaled only the probability of reinforcement operating on a trial) whereas we propose that samples in MTS can serve two functions (i.e., they can signal different probabilities of reinforcement *and* switch the discriminative function of comparisons). In order that their auditory stimuli served that second function, one sound would need to signal that left-key responses will be reinforced after the long sample and right-key responses will be reinforced after the short sample, and the other sound would need to signal that left responses will be reinforced after the short sample and that right responses will be reinforced after the long sample. Although adding this instructional function to the auditory stimuli is likely to increase the number of errors a subject makes in the task (because imperfectly discriminating the auditory stimuli will mean sometimes responding in accordance with the wrong three-term contingency), we see no reason why having this second function should reduce the propensity of these stimuli to signal different reinforcer rates and subsequently control different probabilities of observing. That is, if one can establish stimulus control over attending to the discriminative stimuli in a SD task, the samples in a MTS task should also be able to acquire such control and result in meaningful differences between comparison-discrimination accuracies after the two samples.

Nevin et al.'s (2005) model was motivated largely by the failure of existing BDT models to accommodate the reinforcement-signaling effects reported by Nevin et al. (1982) and the reinforcement-scheduling effects reported by others (e.g., Ferster, 1960; McCarthy & Voss, 1995; Mintz, Mourer, & Weinberg, 1966; Nevin, Cumming, & Berryman, 1963; Nevin, Milo, Odum, & Shahan, 2003). Their solution to this shortcoming was to supplement Alsop's (1991) and Davison's (1991) model with the notion that subjects might occasionally fail to attend to the samples and/or the comparisons presented on a trial. They argued that the probability of attending to the two sets of stimuli was a function of the rate of reinforcement that they signaled relative to the overall context in which they occurred (see their Equation 5 and Equation 6). They then showed that equations incorporating free parameters to measure probabilities of attending described various data sets more accurately than earlier models. Thus, whereas Nevin et al. saw overall reinforcement-rate effects on SD and MTS accuracy as a call for adding parameters to a model already having its own, we see these effects as challenging the basic conceptualization of reinforcement contingencies upon which these models are based. That Nevin et al. did not question the appropriateness of a three-term conceptualization of MTS is also evident in their assumption that the probability of attending to the comparisons [$p(A_C)$] will be the same on S_1 and S_2 trials. Specifically, in not allowing $p(A_C)$ to differ after the two samples, they have maintained the assumption that comparison-selection ratios less than or greater than 1.0 reflect some amount of bias for choosing either comparison. Although we believe that their inclusion of attention parameters is a step in the right direction, we also believe that Nevin et al. (2005) did not go far enough with this concept. They should have broken with tradition and considered the effects of different rates of reinforcement for correct S_1 and correct S_2 responses in terms of different probabilities of comparison attending and, consequently, different comparison-discrimination accuracies, on S_1 and S_2 trials.

Finally, as well as being useful experimental preparations, MTS tasks are used frequently by practitioners providing intensive instruction to people with developmental disabilities and/or significant deficits in verbal behavior. Thus, there is a clear applied need for basic research into experiential variables that affect learning and behavior in these tasks. Furthermore, numerous benefits are accrued from the pursuit of quantitative models of behavior in MTS

tasks (see Bjork, 1973; Mazur, 2006; Rodgers, 2010); they offer greater scientific rigor than verbal theories and they are essential for the formulation of independent measures of MTS performance (see Sidman, 1980) that could serve diagnostic functions when teachers struggle to establish high matching accuracies in their students.⁴ However, some fundamental conceptual issues must be addressed before any quantitative model provides useful measures for practitioners. These issues concern the functions served by the various types of stimuli in the task, and precisely how the constituent responses should be defined. These are far from new issues (see Gulliksen & Wolfle, 1938; Lashley, 1938; Nissen, 1950; Spence, 1952), but a consensus over response and stimulus definitions in simultaneous discriminations was never reached (see Mackintosh, 1974) and some researchers (e.g., Sidman, 1978) even concluded that independent definitions of these terms were impossible. The present article contributes to that discussion by showing that the definitions of stimuli and responses offered in a four-term conceptualization of MTS performance (Jones, 2003), together with the notion that samples can acquire stimulus control of observing, predicted more accurately effects of specific stimulus and reinforcement variables than the definitions offered in the three-term conceptualization adopted in BDTs.

Acknowledgments

We thank the University of Auckland Research Committee for their financial contribution toward the cost of the equipment used in this research, and the Commonwealth Medicine division of the University of Massachusetts Medical School and Grant Number HD046666 from the National Institute of Child Health and Human Development (NICHD) for salary support to the first author during manuscript preparation. We also thank Karen Lionello-DeNolf, Peter Killeen and Peter Urcioli for their thoughtful comments on these data and our interpretation of them, Mick Sibley for his care of the subjects, and Yolande Dunn and other students for their help running this experiment. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of NICHD.

Appendix A

⁴Such measures include an estimate of the stimulus control exerted by samples, the stimulus control exerted by comparisons, and the control exerted by motivational variables such as reinforcement history.

Numbers of responses made by each subject in each condition to C_1 on S_1 trials when C_1 appeared on the left key ($B_{L/S_1 \& C_1}$) and the right key ($B_{R/S_1 \& C_1}$), to C_2 on S_1 trials when C_2 appeared on the left key ($B_{L/S_1 \& C_2}$) and the right key ($B_{R/S_1 \& C_2}$), to C_1 on S_2 trials when C_1 appeared on the left key ($B_{L/S_2 \& C_1}$) and the right key ($B_{R/S_2 \& C_1}$), and to C_2 on S_2 trials when C_2 appeared on the left key ($B_{L/S_2 \& C_2}$) and the right key ($B_{R/S_2 \& C_2}$).

Subject	Cond	$B_{L/S_1 \& C_1}$	$B_{R/S_1 \& C_1}$	$B_{L/S_1 \& C_2}$	$B_{R/S_1 \& C_2}$	$B_{L/S_2 \& C_1}$	$B_{R/S_2 \& C_1}$	$B_{L/S_2 \& C_2}$	$B_{R/S_2 \& C_2}$	R_{L/S_1}	R_{R/S_1}	R_{L/S_2}	R_{R/S_2}
11	1	30	325	55	420	1	20	384	466	0	0	384	466
	2	280	321	0	42	23	240	12	258	280	321	0	0
	3	65	630	78	631	75	402	345	646	0	0	345	646
	4	331	665	90	455	62	675	68	687	331	665	0	0
	5	107	445	330	629	71	360	343	657	0	0	343	657
	6	304	696	50	370	368	495	204	349	304	696	0	0
	7	31	85	428	500	5	45	458	542	0	0	458	542
	8	453	497	0	28	378	500	10	147	453	497	0	0
	9	32	344	154	474	2	30	482	518	0	0	482	518
	10	298	702	27	468	166	728	72	598	298	702	0	0
	11	40	373	148	510	12	116	452	548	0	0	452	548
	12	423	550	4	116	28	477	6	541	423	550	0	0
	13	1	426	226	600	0	278	354	634	0	0	354	634
	14	453	547	0	31	492	569	0	35	453	547	0	0
	15	70	761	85	748	2	982	6	958	0	0	6	958
12	1	41	464	25	455	0	0	496	504	0	0	496	504
	2	451	499	1	0	6	456	56	458	451	499	0	0
	3	3	495	8	536	56	29	516	484	0	0	516	484
	4	507	493	56	47	7	550	15	580	507	493	0	0
	5	39	354	144	486	33	31	519	481	0	0	519	481
	6	498	502	37	23	202	461	90	344	498	502	0	0
	7	1	247	215	480	0	2	502	498	0	0	502	498
	8	494	456	0	2	204	464	1	249	494	456	0	0
	9	16	389	68	497	0	1	489	511	0	0	489	511
	10	528	472	28	18	22	467	21	527	528	472	0	0
	11	17	530	25	484	3	13	487	513	0	0	487	513
	12	491	509	5	4	7	508	19	517	491	509	0	0
	13	3	406	124	515	0	2	518	482	0	0	518	482

Subject	Cond	$B_{LS1\&C1}$	$B_{RS1\&C1}$	$B_{LS1\&C2}$	$B_{RS1\&C2}$	$B_{LS2\&C1}$	$B_{RS2\&C1}$	$B_{LS2\&C2}$	$B_{RS2\&C2}$	R_{LS1}	R_{RS1}	R_{LS2}	R_{RS2}
13	14	491	509	0	1	460	508	1	7	491	509	0	0
	15	62	628	86	628	168	215	466	534	0	0	466	534
	1	22	391	66	484	12	5	423	485	0	0	423	485
	2	398	432	11	5	44	418	19	385	398	432	0	0
	3	33	481	50	452	112	189	333	405	0	0	333	405
	4	348	652	52	356	53	651	35	672	348	652	0	1
	5	53	436	186	675	48	296	325	603	0	0	325	603
	6	302	609	40	382	281	529	75	383	302	609	0	0
	7	5	43	441	454	7	10	469	439	0	0	469	439
	8	350	367	5	9	311	322	5	35	350	367	0	0
	9	2	343	139	439	5	15	485	459	0	0	485	459
	10	361	615	75	352	165	628	67	493	361	615	0	0
	11	57	244	244	433	23	75	431	464	0	0	431	464
	12	371	629	7	254	249	481	139	363	371	629	0	0
14	13	26	146	432	575	16	135	433	532	0	0	433	532
	14	432	428	3	10	428	390	6	9	432	428	0	0
	15	513	301	535	336	29	823	56	790	0	0	56	790
	1	70	524	6	408	3	2	476	524	0	0	476	524
	2	471	471	0	1	8	358	88	460	471	471	0	0
	3	18	538	13	540	47	83	504	496	0	0	504	496
	4	413	438	67	83	11	492	24	488	413	438	0	0
	5	107	252	335	456	89	56	482	518	0	0	482	518
	6	472	528	67	119	438	423	181	205	472	528	0	0
	7	51	20	492	468	3	0	508	492	0	0	508	492
	8	445	459	3	4	290	401	39	154	445	459	0	0
	9	108	379	92	337	3	1	464	475	0	0	464	475
	10	528	472	66	48	39	567	38	524	528	472	0	0
	11	11	482	10	541	12	6	491	509	0	0	491	509
	12	469	515	13	6	7	462	61	512	469	515	0	0
	13	23	229	270	536	9	13	478	484	0	0	478	484
	14	506	494	1	1	503	524	3	13	506	494	0	0

Subject	Cond	$B_{L/S1\&C1}$	$B_{RS1\&C1}$	$B_{L/S1\&C2}$	$B_{RS1\&C2}$	$B_{L/S2\&C1}$	$B_{RS2\&C1}$	$B_{L/S2\&C2}$	$B_{RS2\&C2}$	$R_{L/S1}$	R_{RS1}	$R_{L/S2}$	R_{RS2}
	15	65	738	43	698	300	205	547	453	0	0	547	453

Note. The numbers of reinforcers obtained for each of the four types of correct responses in each condition (i.e., R_L/S_1 , R_R/S_1 , R_L/S_2 and R_R/S_2) are also given. These data represent response and reinforcer tallies summed over the final 20 sessions of each condition.

References

- Alsop, BL. Behavioral models of signal detection and detection models of choice. In: Commons, ML.; Nevin, JA.; Davison, MC., editors. *Signal detection: Mechanisms, models, and applications*. Hillsdale, NJ: Erlbaum; 1991. p. 39-55.
- Alsop BL, Jones BM. Reinforcer control by comparison-stimulus color and location in a delayed matching-to-sample task. *Journal of the Experimental Analysis of Behavior*. 2008; 89:311–331. [PubMed: 18540217]
- Baum WM. On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*. 1974; 22:231–242. [PubMed: 16811782]
- Bjork RA. Why mathematical models? *American Psychologist*. 1973; 28:426–433.
- Blough DS. Error factors in pigeon discrimination and delayed matching. *Journal of Experimental Psychology: Animal Behavior Processes*. 1996; 22:118–131.
- Brown GS, White KG. Measuring discriminability when there are multiple sources of bias. *Behavior Research Methods*. 2009; 41:75–84. [PubMed: 19182126]
- Carter DE, Werner TJ. Complex learning and information processing by pigeons: A critical analysis. *Journal of the Experimental Analysis of Behavior*. 1978; 29:565–601. [PubMed: 16812079]
- Cumming, WW.; Berryman, R. The complex discriminated operant: Studies on matching-to-sample and related problems. In: Mostofsky, DI., editor. *Stimulus generalisation*. Stanford CA: Stanford University Press; 1965. p. 284-330.
- Davison, M. Stimulus discriminability, contingency discriminability, and complex stimulus control. In: Commons, ML.; Nevin, JA.; Davison, MC., editors. *Signal detection: Mechanisms, models, and applications*. Hillsdale, NJ: Erlbaum; 1991. p. 39-55.
- Davison M, Jenkins PE. Stimulus discriminability, contingency discriminability, and schedule performance. *Animal Learning and Behavior*. 1985; 13:77–84.
- Davison M, Nevin JA. Stimuli, reinforcers and behavior: An integration. *Journal of the Experimental Analysis of Behavior*. 1999; 71:439–482. [PubMed: 16812904]
- Davison MC, Tustin RD. The relation between the generalized matching law and signal detection theory. *Journal of the Experimental Analysis of Behavior*. 1978; 29:331–336. [PubMed: 16812059]
- DeMarse TB, Urcuioli PJ. Enhancement of matching acquisition by differential comparison-outcome associations. *Journal of Experimental Psychology: Animal Behavior Processes*. 1993; 19:317–326.
- Ferguson, GA. *Statistical analysis in psychology and education*. New York: McGraw-Hill; 1966.
- Ferster CB. Intermittent reinforcement of matching-to-sample in the pigeon. *Journal of the Experimental Analysis of Behavior*. 1960; 3:259–272. [PubMed: 13822532]
- Gibbon, J.; Church, RM. Sources of variance in an information processing theory of timing. In: Roitblat, HL.; Bever, TG.; Terrace, HS., editors. *Animal cognition*. Hillsdale, NJ: Erlbaum; 1984. p. 465-488.
- Godfrey R, Davison M. Effects of varying sample- and choice-stimulus disparity on symbolic matching-to-sample performance. *Journal of the Experimental Analysis of Behavior*. 1998; 69:311–326. [PubMed: 16812879]
- Grant DS. Symmetrical and asymmetrical coding of food and no-food samples in delayed matching in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*. 1991; 17:186–193.
- Green, DM.; Swets, JA. *Signal-detection theory and psychophysics*. New York: Wiley; 1966.
- Gulliksen H, Wolfle DA. A theory of learning and transfer. *Psychometrika*. 1938; 3:127–149.
- Harnett P, McCarthy D, Davison M. Delayed signal detection, differential reinforcement and short-term memory in the pigeon. *Journal of the Experimental Analysis of Behavior*. 1984; 42:87–111. [PubMed: 6481301]
- Hautus MJ. Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instrumentation and Computers*. 1995; 27:46–51.
- Heinemann EG, Chase S, Mandell C. Discriminative control of “attention.”. *Science*. 1968; 160:553–554. [PubMed: 5689414]

- Honig, WK. Attention and the modulation of stimulus control. In: Mostofsky, D., editor. *Attention: Contemporary studies and analyses*. New York: Appleton-Century-Crofts; 1969.
- Jackson WJ, Pegram GV. Comparison of intra- vs. extradimensional transfer of matching by rhesus monkeys. *Psychonomic Science*. 1970; 19:162–163.
- Jones BM. Quantitative analyses of matching to sample performance. *Journal of the Experimental Analysis of Behavior*. 2003; 79:323–350. [PubMed: 12908761]
- Jones BM, Davison MC. Reporting contingencies of reinforcement in concurrent schedules. *Journal of the Experimental Analysis of Behavior*. 1998; 69:161–183. [PubMed: 16812872]
- Jones BM, White KG. Stimulus discriminability and sensitivity to reinforcement in delayed matching-to-sample. *Journal of the Experimental Analysis of Behavior*. 1992; 58:159–172. [PubMed: 16812662]
- Kangas BD, Branch MN. Empirical validation of a procedure to correct position and stimulus biases in matching-to-sample. *Journal of the Experimental Analysis of Behavior*. 2008; 90:103–112. [PubMed: 18683615]
- Katz JL. Two types of bias in psychophysical detection and recognition procedures: Nonparametric indices and effects of drugs. *Psychopharmacology*. 1989; 97:202–205. [PubMed: 2498929]
- Lashley KS. Conditional reactions in the rat. *Journal of Psychology*. 1938; 6:311–324.
- Mackintosh, N. *The psychology of animal learning*. New York: Academic Press; 1974.
- Maki WS Jr, Gillund G, Hague G, Siders WA. Matching to sample after extinction of observing responses. *Journal of Experimental Psychology: Animal Behavior Processes*. 1977; 3:285–296.
- Mazur JE. Mathematical models and the experimental analysis of behavior. *Journal of the Experimental Analysis of Behavior*. 2006; 85:275–291. [PubMed: 16673829]
- McCarthy D, Davison M. Signal probability, reinforcement, and signal detection. *Journal of the Experimental Analysis of Behavior*. 1979; 32:373–386. [PubMed: 512570]
- McCarthy DC, Davison MC. The interaction between stimulus and reinforcer control on remembering. *Journal of the Experimental Analysis of Behavior*. 1991; 56:51–66. [PubMed: 1940763]
- McCarthy DC, Voss P. Delayed matching-to-sample performance: Effects of relative reinforcer frequency and of signaled versus unsignaled reinforcer magnitudes. *Journal of the Experimental Analysis of Behavior*. 1995; 63:31–51.
- Mintz DE, Mourer DJ, Weinberg LS. Stimulus control in fixed-ratio matching to sample. *Journal of the Experimental Analysis of Behavior*. 1966; 9:627–630. [PubMed: 5970383]
- Nevin JA. Signal detection theory and operant behavior: A review of David M. Green and John A. Swets' *Signal Detection Theory and Psychophysics*. *Journal of the Experimental Analysis of Behavior*. 1969; 12:475–480.
- Nevin JA, Cumming WW, Berryman R. Ratio reinforcement of matching behavior. *Journal of the Experimental Analysis of Behavior*. 1963; 6:149–154. [PubMed: 13938218]
- Nevin JA, Davison M, Shahan TA. A theory of attending and reinforcement in conditional discriminations. *Journal of the Experimental Analysis of Behavior*. 2005; 84:281–303. [PubMed: 16262190]
- Nevin JA, Grosch J. Effects of signaled reinforcer magnitude on delayed matching-to-sample performance. *Journal of Experimental Psychology: Animal Behavior Processes*. 1990; 16:298–305.
- Nevin JA, Jenkins P, Whittaker SG, Yarensky P. Reinforcement contingencies and signal detection. *Journal of the Experimental Analysis of Behavior*. 1982; 37:65–79. [PubMed: 7057131]
- Nevin JA, Milo J, Odum AL, Shahan TA. Accuracy of discrimination, rate of responding, and resistance to change. *Journal of the Experimental Analysis of Behavior*. 2003; 79:307–321. [PubMed: 12908760]
- Nissen HW. Description of the learned response in discrimination behavior. *Psychological Review*. 1950; 57:121–131.
- Ray BA. Selective attention: The effects of combining stimuli which control incompatible behavior. *Journal of the Experimental Analysis of Behavior*. 1969; 12:539–550. [PubMed: 16811373]
- Rodgers JL. The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*. 2010; 65(1):1–12. [PubMed: 20063905]

- Sidman M. Remarks. *Behaviorism*. 1978; 6:265–268.
- Sidman M. A note on the measurement of conditional discrimination. *Journal of the Experimental Analysis of Behavior*. 1980; 33:285–289. [PubMed: 7365409]
- Sidman, M. Functional analysis of emergent verbal classes. In: Thompson, T.; Zeiler, MD., editors. *Analysis and integration of behavioral units*. Hillsdale, NJ: Erlbaum; 1986. p. 213-245.
- Sidman M. Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*. 2000; 74:127–146. [PubMed: 10966100]
- Spence KW. The nature of the response in discrimination learning. *Psychological Review*. 1952; 59:89–93. [PubMed: 14912194]
- Spetch ML, Wilkie DM. A systematic bias in pigeons' memory for food and light durations. *Behaviour Analysis Letters*. 1982; 2:267–274.
- Wixted JT. A signal detection analysis of memory for nonoccurrence in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*. 1993; 19:400–411.
- Wixted JT, Gaitan SC. Stimulus salience and asymmetric forgetting in the pigeon. *Learning and Behavior*. 2004; 32(2):173–182. [PubMed: 15281389]
- Wright AA, Sands SF. A model of detection and decision processes during matching to sample by pigeons: Performance with 88 different wavelengths in delayed and simultaneous matching tasks. *Journal of Experimental Psychology: Animal Behavior Processes*. 1981; 7:191–216.

Fig. 1.

The proportion of trials on which a correct response was made plotted as a function of the disparity between the sample stimuli for each level of comparison-stimulus disparity. Unfilled circles represent data from conditions that were systematic replications of other conditions and so have not been joined by lines.

Fig. 2.

Point estimates of stimulus discriminability (log d , Equation 1) and response bias (log b , Equation 2) from Davison and Tustin's (1978) model plotted as a function of sample-stimulus disparity for each level of comparison-stimulus disparity (top panel) and as a function of comparison disparity for each level of sample disparity (bottom panel). These data represent the mean estimates across individual subjects. Thus, standard-error bars have been shown around each data point.

Fig. 3.

Measures of the accuracy with which comparison stimuli were discriminated on S_1 and S_2 trials, separately calculated using Equations 3 and 4 respectively, and plotted for CRF and Ext trials as a function of sample-stimulus disparity for each level of comparison-stimulus disparity. These measures are afforded by Jones' (2003) conceptualization of reinforcement contingencies in MTS tasks, offer estimates of the degree of stimulus control exerted by the relevant dimension of the comparison stimuli (e.g., their colors) over responding at the test phase of each trial following one sample type, and are obtained by calculating the degree to which the left/right response ratio differs on the two comparison configurations after one sample (e.g., $C_1-S_1-C_2$ and $C_2-S_1-C_1$). Negative log D values indicate that a subject made more errors than correct responses.

Fig. 4.

Measures of the position (or location) bias in each subject's responding at the comparison-presentation phase of a trial on trials involving reinforcement of all correct responses (filled circles) and extinction for those responses (open circles) plotted as a function of sample-stimulus disparity for each level of comparison-stimulus disparity. Position bias has been calculated by taking the absolute values of log left over right response ratios, and constitutes response bias in Jones' (2003) theory.

Fig. 5.

Frequency distributions of the latencies of Pigeon 12 to peck a comparison-stimulus key on CRF (solid lines) and Ext trials (dotted lines). Latencies were allocated to bins of 0.02 s before the proportion of all responses that fell in specific bins was calculated. The right-most bin shows the proportion of responses that exceeded a latency of 4 s. Rows of graphs show data from conditions that arranged the same sample disparity but various comparison disparities. Columns of graphs show data from conditions that arranged the same comparison disparity but various sample disparities.

Table 1

The sequence of conditions, the probabilities used to generate the random-dot patterns depicting the sample stimuli (S_1 and S_2) and the comparison stimuli (C_1 and C_2), indices of sample disparity and comparison disparity, the comparison stimulus associated with reinforcement (Positive Comp.), and the number of training sessions in each condition.

Cond	Pixel Probability				Sample Disparity	Comp. Disparity	Positive Comp.	No. of Sessions
	S_1	S_2	C_1	C_2				
1	.3	.7	.3	.7	.4	.4	C_2	44
2	.3	.7	.3	.7	.4	.4	C_1	42
3	.3	.7	.45	.55	.4	.1	C_2	41
4	.3	.7	.45	.55	.4	.1	C_1	42
5	.45	.55	.45	.55	.1	.1	C_2	40
6	.45	.55	.45	.55	.1	.1	C_1	40
7	.45	.55	.3	.7	.1	.4	C_2	40
8	.45	.55	.3	.7	.1	.4	C_1	45
9	.4	.6	.3	.7	.2	.4	C_2	41
10	.4	.6	.45	.55	.2	.1	C_1	43
11	.4	.6	.4	.6	.2	.2	C_2	54
12	.3	.7	.4	.6	.4	.2	C_1	54
13	.45	.55	.4	.6	.1	.2	C_2	40
14	.47	.53	.3	.7	.06	.4	C_1	49
15	.3	.7	.47	.53	.4	.06	C_2	44

The results obtained when Ferguson's (1966) non-parametric test for monotonic trends was applied in analyses of proportion correct scores (Figure 1), point estimates of $\log d$ and $\log b$ from the Davison and Tustin (1978) model (Figure 2), comparison-discrimination accuracies ($\log D$) on reinforcement and extinction trials (Figure 3), and position biases ($\log B_L/B_R$) on reinforcement and extinction trials (Figure 4) as a function of sample and comparison disparities.

Table 2

Analysis	Figure	DV	IV	SD_y/CD_y	Increase/ Decrease	k	ΣS	z score
1	1	Propn. Corr	SD_y	$CD_y = .4$	decrease	4	22	3.57*
2				$CD_y = .2$	decrease	3	8	1.83*
3				$CD_y = .1$	decrease	3	10	2.35*
4			CD_y	$SD_y = .4$	decrease	4	20	3.23*
5				$SD_y = .2$	decrease	3	8	1.83*
6				$SD_y = .1$	none	3	-2	0.26
7	2	$\log d$	SD_y	$CD_y = .4$	decrease	4	24	3.91*
8				$CD_y = .2$	none	3	6	1.31
9				$CD_y = .1$	decrease	3	8	1.83*
10			CD_y	$SD_y = .4$	decrease	4	24	3.91*
11				$SD_y = .2$	decrease	3	12	2.87*
12				$SD_y = .1$	decrease	3	8	1.83*
13	2	$\log b$	SD_y	$CD_y = .4$	increase	4	18	2.87*
14				$CD_y = .2$	none	3	6	1.31
15				$CD_y = .1$	increase	3	12	2.87*
16			CD_y	$SD_y = .4$	decrease	4	24	3.91*
17				$SD_y = .2$	decrease	3	12	2.87*
18				$SD_y = .1$	decrease	3	12	2.87*
19	3	$\log D$ (CRF trials)	SD_y	$CD_y = .4$	none	4	-3	0.34
20				$CD_y = .2$	none	3	2	0.26
21				$CD_y = .1$	none	3	-4	0.78
22			CD_y	$SD_y = .4$	decrease	4	24	3.91*

Analysis	Figure	DV	IV	SDy/CDy	Increase/ Decrease	k	ΣS	z score
23				SDy=.2	decrease	3	12	2.87*
24				SDy=.1	decrease	3	11	2.61*
25	3	log D (Ext trials)	SDy	CDy=.4	decrease	4	24	3.91*
26				CDy=.2	decrease	3	12	2.87*
27				CDy=.1	decrease	3	10	2.35*
28			CDy	SDy=.4	increase	4	21	3.40*
29				SDy=.2	increase	3	12	2.87*
30				SDy=.1	increase	3	10	2.35*
31	4	log B _L /B _R (CRF trials)	SDy	CDy=.4	none	4	4	0.51
32				CDy=.2	none	3	2	0.26
33				CDy=.1	none	3	0	0.26
34			CDy	SDy=.4	decrease	4	16	2.55*
35				SDy=.2	decrease	3	12	2.87*
36				SDy=.1	none	3	4	0.78
37	4	log B _L /B _R (Ext. trials)	SDy	CDy=.4	decrease	4	24	3.91*
38				CDy=.2	decrease	3	8	1.83*
39				CDy=.1	decrease	3	12	2.87*
40			CDy	SDy=.4	none	4	2	0.17
41				SDy=.2	none	3	4	0.78
42				SDy=.1	none	3	6	1.31

Note. The number of subjects contributing data to each test (N) was always 4. DV refers to the dependent variable used in the analysis, IV refers to the independent variable in the analysis, ΣS refers to the S statistic summed across subjects, k refers to the number of conditions used in the analysis, SDy refers to the sample disparity arranged in the set of conditions analyzed, and CDy refers to the comparison disparity arranged. Data in the column headed Increase/Decrease indicate whether the DV increased or decreased as the IV (either sample or comparison disparity) decreased. Cases of tied data were counted as zero in the calculation of S for each subject. All tests were 1-tailed because a directional effect was hypothesized.

* $p < .05$