THE EFFECTS OF THE DELAY IN A DELAYED MATCH-TO-SAMPLE PROCEDURE ON ACQUISITION AND TRANSFER

Kimberly N. Smith, B.S.

Thesis Prepared of the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

December 2005

APPROVED:

Manish Vaidya, Major Professor
Sigrid Glenn, Committee Member
Richard Smith, Committee Member and Chair of the Department of Behavior Analysis
David Hartman, Dean of the College of Public Affairs and Community Service
Sandra L. Terrell, Dean of the Robert B. Toulouse School of Graduate Studies
Smith, Kimberly N.  The Effects of the Delay in a Delayed Match-To-Sample Procedure on Acquisition and Transfer. Master of Science (Behavioral Analysis), December 2005, 33 pp., 2 tables, 9 illustrations, references, 25 titles.

Twenty-six participants, divided into three groups, learned to relate English words to Czech and Portuguese words in a matching-to-sample procedure. Half the word pairs were learned using English words as samples and foreign words as corresponding comparisons and the other half were learned with the foreign words serving as samples and English words as corresponding comparisons. The only difference in training across the three groups involved a programmed delay between the removal of the sample stimulus and the presentation of comparison stimuli. For Group 0, Group 2, and Group 8, the programmed delay values between sample offset and comparison onset were 0 s, 2 s, and 8 s, respectively. Test trials assessed the extent to which the conditional discriminations established during training had become reversible or the extent to which the effects of learning had transferred to a new situation. The results suggest that the likelihood of transfer was greatest for the group that learned the task with the largest delay (i.e., an 8 s delay between sample offset and comparison onset).
ACKNOWLEDGEMENTS

I would like to thank Dr. Manish Vaidya for exposing me to the world of basic science, for his encouragement, guidance, and supervision, Dr. Sigrid Glenn for inspiring me with her hard work and dedication and taking an interest in my research, and Dr. Richard Smith for years of advice, guidance, and giving me the opportunity to work in his laboratory.

I would like to recognize the other individuals who made this project possible and provided assistance along the way. Molly Meadows for helping me run sessions, Melissa Morehead and Nicole Dorey for reading some initial drafts and for their support, Carla Smith for her help and friendship while working on this degree, and my parents for their various support throughout the years. And my boyfriend, Jason Frame, for his continuous encouragement, willingness to move, and all the other sacrifices he has made for my education.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>v</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>METHOD</td>
<td>7</td>
</tr>
<tr>
<td>Subjects</td>
<td></td>
</tr>
<tr>
<td>Setting and Apparatus</td>
<td></td>
</tr>
<tr>
<td>Procedure</td>
<td></td>
</tr>
<tr>
<td>RESULTS</td>
<td>11</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>15</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>31</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1.
Word Pairs Used During Baseline and Training .................................................. 19

Table 2.
Word Pairs Used During Testing........................................................................ 20
LIST OF ILLUSTRATIONS

Figure 1.
Percent correct in baseline probes for English and foreign word samples for each subject for Groups 0, 2, and 8. Asterisk denotes 0% for those trials................................................................. 22

Figure 2.
Percent correct across blocks during training condition for Group 0.........................23

Figure 3.
Percent correct across blocks during training condition for Group 2.........................24

Figure 4.
Percent correct across blocks during training condition for Group 8.........................25

Figure 5.
Average latency in seconds for sample and comparison stimuli during the training condition for each group. Sample and comparison latencies are further divided into English and foreign word stimuli.................................................................26

Figure 6.
Percent correct across blocks during testing condition for Group 0. The dotted line represents 80%. .................................................................27

Figure 7.
Percent correct across blocks during testing condition for Group 2. The dotted line represents 80%.................................................................28

Figure 8.
Percent correct across blocks during testing condition for Group 8. The dotted line represents 80%.................................................................29

Figure 9.
Average latency in seconds for sample and comparison stimuli during the testing condition for each group. Sample and comparison latencies are further divided into English and foreign word stimuli.................................................................30
INTRODUCTION

Stimulus equivalence refers to the observation that training a few overlapping conditional
discriminations in verbally sophisticated subjects results in a host of other systematic conditional
discriminations that were not directly trained (Sidman & Tailby, 1982). For example, a subject
may be taught to select the word dog (Stimulus B) from an array of English words in the
presence of a picture of a dog (Stimulus A) and to select the written word perro (Stimulus C)
from an array of Spanish words in the presence of the written word dog. (This type of training
will, hereafter, be designated as A-B and B-C, where the first term refers to the sample stimulus
and the second term refers to the comparison stimulus in the conditional discrimination
procedure). Having learned A-B and B-C conditional discriminations, verbally sophisticated
subjects behave in systematic and predictable ways on probe trials in which the trained
conditional discriminations are reversed or intermixed. For example, subjects will, without
explicit training, select (i) picture of dog given dog (B-A), (ii) dog given perro (C-B), (iii) perro
given the picture of a dog (A-C), and (iv) the picture of a dog given perro (C-A). These
outcomes are said to reflect the formation of an equivalence relation and the stimuli are said to
comprise an equivalence class (Sidman & Tailby, 1982).

Although the outcomes described above have mainly been documented in the
laboratory, they appear to closely mimic some important features of languages. For example,
the complete substitutability of the three words used in the description above in the context of a
conditional discrimination is not unlike a vocabulary, albeit a rudimentary one, in which the
words dog and perro and the picture of the dog refer to each other. In this sense, equivalence
relations are thought to be laboratory models of referential meaning. In addition, results from
other studies show that a behavioral function established for one member of an equivalence
class will extend to other members, again without direct training. For example, if a subject was
taught to say “woof woof” in the presence of the picture of the dog, he or she will also be more
likely to say “woof woof” in the presence of the written words dog and perro (Sidman, 1994;
Roche, Barnes-Holmes, Barnes-Holmes, Stewart, & O’Hora, 2002). These data have suggested to many that laboratory research on equivalence relations can model several critical aspects of natural languages. Indeed, some researchers have recently offered a complete reinterpretation of verbal learning in terms of equivalence and equivalence-like dynamics (Roche et al., 2002).

Despite the sweeping interpretations of language and other complex phenomena offered in terms of equivalence or relational framing (e.g., Barnes-Holmes & Hayes, 2003; Sidman, 2000), there is much that remains to be understood about the phenomena that characterize stimulus equivalence. For example, although the baseline conditional discriminations are considered to be the putative basis for the development of equivalence relations the relation between the two is not well understood. Is the establishment of a conditional discrimination via direct reinforcement a sufficient condition for the development of equivalence relations? Results from a large number of studies show a lag between the acquisition of the baseline conditional discriminations and the development (or emergence) of equivalence relations (e.g., Sidman, Kirk, & Willson-Morris, 1985; Sidman & Tailby, 1982; Spradlin, Cotter, & Baxley, 1973). These data suggest that acquisition of baseline conditional discriminations in not a sufficient condition for the development of equivalence-consistent responding. For example, the baseline conditional discriminations may have to be over-trained before equivalence-consistent responding becomes likely.

Alternatively, is the establishment of a conditional discrimination via direct reinforcement a necessary condition? Data from several studies suggests that the answer is no. For example, Harrison and Green (1990) found that subjects could learn to behave in systematic ways in a conditional discrimination procedure with no programmed contingencies. Not only did subjects learn certain conditional relations, but the authors also showed that the learned conditional relations were reversible and recombinative, consistent with the defining characteristics of stimulus equivalence classes. In another study, Tonneau and González (2004) showed that simple exposure to stimulus pairings – with no response requirement of any
sort -- was sufficient to establish equivalence-consistent responding in adult humans. Taken together, these data suggest that much remains to be understood about the conditions under which trained conditional discriminations become equivalence relations.

It appears that the manner in which baseline relations are trained affects the likelihood that equivalence relations will emerge. For example, procedures in which the subject is not required to touch the sample stimulus produce slightly lower overall accuracy on equivalence tests relative to preparations in which an observing response is required (Fields & Verhave, 1987). Others have reported that certain training structures, referring to the order and type of overlapping conditional discrimination training, are more effective than others in producing equivalence consistent outcomes (e.g., Arntzen & Holth, 2000; see also Saunders & Green, 1999).

One way to order all of these findings is to consider equivalence consistent outcomes as a kind of transfer of learning. The training conditions in equivalence experiments serve to establish conditional and discriminative stimulus functions for stimuli. The testing conditions provide an occasion upon which the trained stimulus functions are transferred or extended to other stimuli. For example, the stimuli A and B can be said to acquire conditional and discriminative stimulus functions, respectively, as a result of training to select B given A. This performance is said to be symmetric when the trained conditional functions for A are seen to extend to B and the simple discriminative functions of B are said to extend to A. In these terms, all equivalence relations -- symmetry, transitivity, and symmetrical-transitivity -- reflect only the transfer or extension of directly established conditional and discriminative stimulus functions to other stimuli involved in the training and testing conditions. From this perspective, questions about the conditions under which equivalence classes develop are really questions about the conditions under which learning will transfer.

Transfer of Learning
The issue of transfer of learning has been of interest to behavioral and non-behavioral psychologists alike. For example, researchers in the general experimental psychology tradition have defined transfer of learning, as occurring “whenever prior-learned knowledges and skills affect the way in which new knowledges and skills are learned and performed” (Cormier & Hagman, 1987, p. 1). Gick and Holyoak (1987) provide a more generic definition of transfer of knowledge: “transfer is a phenomenon involving change in the performance of a task as a result of the prior performance of a different task” (p. 10). Cormier and Hagman (1987) make a further distinction between training that focuses on transfer of learned skills and training that focuses on acquisition of said skills. Training that focuses on transfer may maximize observed transfer; however, it may do so at increased costs (i.e., increased session duration, noticeable decrements in accuracy, etc.). This kind of training is designed to prepare the learner to perform not only the “original task, but [it is] also preparing them for related but distinct target tasks” (p.3). Training for acquisition, on the other hand, assumes that the training materials are the same for training and transfer tasks. In general, trainers and educators often base their evaluations on the trainee’s performance during acquisition, while transfer and/or maintenance of relations is sometimes entirely ignored (Cormier & Hagman, 1987; Schmidt & Bjork, 1992).

The relation between conditions of acquisition and the likelihood of transfer has received experimental attention, especially since the 1980’s. Research topics have included concept formation, motor skills, verbal learning tasks, matching to sample, and stimulus equivalence (Schmidt & Bjork, 1992). For example, Schroth (1995) investigated the likelihood of transfer by manipulating the delay of feedback. Some subjects received fixed feedback delivered after a delay of 0 s, 10 s, or 30 s and others received delayed feedback delivered randomly after 10 s or 30 s. Subjects were required to learn a concept that varied on several dimensions, including color (red, green, and blue), size (1.27 cm, 2.45 cm, and 3.81 cm), and form (X, H, O). Two dimensions of the stimuli were relevant to the concept and one dimension was not (which dimensions were relevant varied across concepts trained). An example of a concept would be
Schroth (1997) examined the effects of varied training conditions on acquisition and maintenance. The stimuli and concepts are described above. A selective or receptive training procedure was used to teach the concepts; delayed feedback was parametrically manipulated across training conditions (0 s, 15 s, or 30 s). Testing for transfer occurred immediately after training with a retention test occurring 7 days later. Immediate and delayed testing are described above. Subjects in the most difficult conditions—that is, receptive training with 30 s delayed feedback—did the best in both the immediate and delayed testing conditions. The combined results of Schroth’s studies show degradation in accuracy during acquisition when the training conditions were difficult compared to easier training conditions. However, subjects were more accurate in the transfer condition when the training condition was difficult compared to easier training conditions.

Bjork and Allen (1970) conducted a verbal learning experiment where they presented one or two trigrams or “three common four-letter nouns” (p. 568) per trial. The easy and difficult conditions began the same way. A trigram was presented followed by an easy task (reading a series of three or five digit numbers at a rate of 1.5 s) after the easy task the trigram was presented for an additional 2.5 s. The subjects then engaged in either an easy task or a difficult task (reading a series of 8 or 20 four-digit numbers) before having to recall the previously seen trigram. If a difficult task was performed before a subject was asked to recall a trigram they performed slightly but consistently better in the recall task, with the average percent of errors being lower than if an easy task was performed before a subject was asked to recall a trigram.

Researchers working within the behavioral tradition have noted similar relations between acquisition and transfer. For example, Wright, Rivera, Cook, Sands, and Delius (1988) trained identity matching to two groups of pigeons with color cartoons. Group 152 was presented with
152 different stimuli per session each day using a trial unique procedure. Group 2 was presented with 2 stimuli per session each day. Wright et al. found that Group 2 reached the acquisition criterion approximately ten times faster than the Group 152. Transfer was assessed to see if concept learning had been established using novel stimuli. Group 152 was considerably more accurate in the transfer test than Group 2.

The results of these and other studies can be summarized in brief: increasing the difficulty by manipulating task-relevant variables appears to enhance transfer to novel tasks even as it retards acquisition in the original training task. More difficult training conditions may affect the way a subject learns the material, forcing a correspondence between properties of the stimuli that experimenters hope will control the subjects' behavior and those that actually do (e.g., Horne & Lowe, 1996; McIlvane & Dube, 2003). The purpose of the present experiment was to manipulate the delay in a delayed match to sample procedure in a verbal learning task and observe the effects on performance during acquisition and transfer. Subjects matched English-to-foreign words and foreign-to-English word and then tested for symmetry in a delayed match-to-sample procedure. The delay between the sample offset and comparison onset was parametrically manipulated across conditions. After subjects learned the word pairs, degree of transfer was measured by assessing the extent to which the trained stimulus function had become reversible using a symmetry test.
METHOD

Subjects

Twenty-six students attending the University of North Texas (10 male and 16 female) participated. Participants were recruited via advertisements in the university paper or flyers posted in a variety of locations around campus and were randomly assigned to one of three groups (described below). Applicants were excluded from participating if they were familiar with behavior analysis or with Spanish, Portuguese, or Czech languages. The only criterion for participation was meeting the constraints described above and availability at the time sessions were scheduled. Participants received $0.05 for every correct response during training and $1 for each session, which totaled to a minimum of $6 per hour. Some participants received an additional $5.00 at the end of their participation to ensure that the subjects earned a minimum of $6.00 per hour. For all participants, payment occurred after the participant was debriefed. There were eight, eight, and ten subjects in the 0 s, 2 s, and 8 s groups, respectively.

Setting and Apparatus

Sessions were conducted in a 2 m by 3 m room equipped with a chair and a small table. The apparatus sat atop the table and comprised a Macintosh Power Book (G3) enclosed in a touch screen adapter (Troll Touch, Inc.). A custom-written program (MTS v. 11.6.7, see also Dube & Hirris, 1991) was used to present stimuli, detect responses, manage contingencies and collect data. In order to minimize contact with the keyboard, a piece of thin cardstock (from a file folder) was used to cover the keyboard. The room also contained another table, chair, and a personal computer, which remained unplugged throughout the experiment.

Procedure

Sessions consisted of three distinct phases in which (a) preexisting relations among the words were measured; (b) some conditional discriminations among words were trained; and (c) untrained conditional discriminations were tested. In most cases, the three phases were
conducted in one session. If session duration exceeded two hours, the session was terminated and the participant was excluded from the study. This happened with one subject (S24).

Baseline. The experimenter directed the participant to the workroom and asking him/her to sit in the chair in front of the computer. The experimenter then read the following instructions aloud: “The computer has a touch screen. Please do not touch the keyboard. The program will tell you when the session is over, and it is your job to figure out what to do.” Following this, subjects were exposed to one 8-trial block in which each conditional discrimination to be trained was presented once without programmed feedback for participants’ choices. A trial began with the presentation of a word in the middle of the screen (hereafter, the sample stimulus). Touching the sample stimulus removed it from the screen and immediately produced four other words in the corners of the screen (hereafter, the comparison stimuli). Touching any of the four comparison stimuli terminated the trial. The inter-trial interval (ITI) was 1.5 s. Four trials presented English words as sample stimuli with foreign words (including the counterpart to the English words) as comparison stimuli and another four trials presented foreign words as samples and English words (including the counterpart to the foreign words) as comparison stimuli (see Table 1). Subjects’ participation was discontinued if the overall accuracy during the baseline condition exceeded 62.5%. No participants were excluded on this basis.

Training. The three experimental groups were distinguishable only in terms of the conditions experienced during the training phase. For subjects in all groups, a trial began with the presentation of the sample stimulus in the middle of the screen. Touching the sample stimulus removed it from the screen and produced four comparison stimuli in the corners of the screen after delays that varied as a function of group membership. Specifically, for S1-S8, touching the sample stimulus removed it from the display and immediately (0 s delay) presented the array of comparison stimuli -- a zero-delay MTS procedure (hereafter, Group 0). For S9-S16, touching the sample stimulus removed it from the display and initiated a 2 s delay during which the screen was dark and after which the array of comparison stimuli was presented.
(hereafter, Group 2). For S17-S26, touching the sample stimulus removed it from the display and initiated an 8 s delay during which the screen was dark and after which the array of comparison stimuli was presented. All other contingencies were identical across groups and participants. For all subjects, touching the experimenter-designated correct comparison stimulus resulted in the word “CORRECT”, enclosed in a box, appearing on the screen for 1 s with two short tones. Incorrect responses resulted in a 3 s timeout during which the screen was dark. All trials had an ITI of 1.5 s regardless of the outcome of the trial.

Participants in all groups were taught eight conditional discriminations. Four conditional discriminations used the English words food, bathroom, farm, and ball as sample stimuli and the foreign words (potrava, koupelna, fazenda, and esfera) as comparison stimuli. The other four conditional discriminations used the foreign words (postel, duum, margem, and comboio) as sample stimuli and the English words (bed, house, bank, and train) as comparison stimuli. Each of the eight conditional discrimination trials was presented once before any could repeat. This training condition continued until the subject responded correctly on 24 consecutive trials or until he/she had been exposed to 384 training trials. In the latter case, the subject was invited to return for a second session that was identical to the first session. Participants who met the acquisition criterion were exposed to testing trials in which stimuli previously functioning as comparison stimuli were presented as sample stimuli and vice versa.

Testing. Trials presented during the testing condition were different from trials presented during the training condition in two ways: first, all feedback for subject’s choices was discontinued and, second, the test trials presented stimuli that appeared as sample stimuli during the training condition as comparison stimuli and vice versa. Specifically, four test trial-types presented the foreign words (potrava, koupelna, fazenda, and esfera) as sample stimuli and English words (food, bathroom, farm and ball) as comparison stimuli. The other four conditional discriminations used the English words (bed, house, bank, and train) as sample stimuli and the foreign words (postel, duum, margem, and comboio) as comparison stimuli (see
Table 2). The testing condition consisted of 160 trials for 25 out of 26 subjects (S1 had 80 testing trials). Each of the 8 test trial-types was presented once before any were allowed to repeat. Subject participation was considered complete after 160 testing trials.

Debriefing

Debriefing occurred in a separate session. Each participant completed a questionnaire and a sorting task, followed by an explanation of the study, their own outcomes, and payment. The sorting task consisted of 16 words printed on paper measuring approximately 3.25" x 1.25"; there was one word per piece of paper. Each participant sorted the 16 words; there were no instructions for sorting.
RESULTS

Figure 1 presents accuracy on baseline probes for each subject in each of the three groups. The asterisk denotes zero percent for that group of trials. The top panel in Figure 1 presents data from Group 0 (S1-S8). This panel shows that the subjects’ accuracy was often around chance levels. The exception was S1’s accuracy on trials involving English words as samples and foreign words as comparisons. Four of the eight subjects’ performance (S1, S2, S3, and S5) was more accurate on trials with English words as samples relative to trials using foreign words as samples. The middle panel presents data from Group 2 comprising eight subjects. This panel shows that accuracy for these subjects was also often around chance levels. The bottom panel presents data for 10 subjects assigned to Group 8. As before, this panel shows that accuracy on the baseline probes fluctuated around chance levels. For two subjects (S20 and S23), accuracy on trials with English words as samples was well above chance.

Figures 2, 3, and 4 present data from the training condition for each subject in Groups 0, 2 and 8, respectively. Figure 2 presents data from the training condition for subjects assigned to Group 0. This figure shows that all subjects acquired the eight conditional discriminations within 12 – 25 blocks (or 96 – 200 trials). One block equals four trials, which are separated into English word samples and foreign word samples. A detailed look at the patterns of acquisition shows that, in general, conditional discriminations involving English and foreign words as samples were acquired at the same rate. There were, however, some exceptions. Two subjects (S3 and S7) acquired the conditional discriminations involving English word samples three to four blocks earlier than conditional discriminations involving foreign word samples. For two other subjects (S1 and S8), accuracy on trials involving foreign word samples was slightly higher relative to trials involving English word samples.

Figure 3 presents data from the training condition for subjects assigned to Group 2. This figure shows that 7 of 8 subjects acquired the conditional discriminations within 10 – 18 blocks
(or 80 – 144 trials). One subject (S16) failed to acquire the conditional discriminations after being exposed to 48 blocks (or 384 trials). This subject was invited to return for a second session. During this session, S16 acquired the conditional discriminations within 13 blocks (or 104 trials). As with Group 0, there were no systematic differences in the rates of acquisition for conditional discriminations involving English or foreign words as samples. However, S10’s performance was generally more accurate on trials involving English words as samples and S16’s performance was slightly more accurate on trials involving foreign words as samples during the second session.

Figure 4 presents data from the training condition for subjects assigned to Group 8. This figure shows that 8 out of 10 subjects acquired the conditional discriminations within 7 – 20 blocks (or 56 – 200 trials). For most subjects, there were no systematic differences in the rates of acquisition between conditional discriminations involving English or foreign words as samples. However, performances of three subjects (S17, S21, and S22) were slightly more accurate on trials involving foreign words as samples relative to trials involving English words as samples. Two subjects’ performances (S20 and S23) were slightly more accurate on trials involving English words as samples relative to trials involving foreign words as samples.

Finally, two subjects, S24 and S25, failed to acquire the conditional discrimination after exposure to 48 and 96 blocks (or 384 and 768 trials), respectively. For S24, accuracy improved slowly but the session was terminated when the pre-established session duration limit was exceeded. S25’s performance consistently fluctuated around chance levels of accuracy during each of two sessions. Detailed analyses of error patterns indicated position biases – the subject selected stimuli occurring in one position on 123 trials during Session 1 and a different position on 131 trials during Session 2. As these subjects did not meet criterion to move to the testing condition, their participation was discontinued following 2 sessions.

Figure 5 presents average latencies from the training condition for each of the three groups. On each graph, the average latencies to respond to the sample stimulus and to select
comparison stimuli are presented separately for trials involving English and foreign words. Mean latencies for Group 0, Group 2, and Group 8 are presented in the top, middle, and bottom panels, respectively. The data presented in the top and middle panels, for subjects in Group 0 and Group 2 respectively, show that latencies were similar for the foreign word samples and the English word samples; there was no systematic difference for Group 0 in the latencies to respond to English or foreign words in an array of comparison stimuli. The data from Group 8 (bottom panel), however, show that subjects spent more time looking at the foreign word sample stimuli relative to English word samples. The figures also show that the latencies to select English word comparisons were slightly higher in Group 2 and 8 than latencies to select foreign word samples.

Figures 6, 7, and 8 present data from the testing condition for each subject in Groups 0, 2 and 8, respectively. Each data point represents percent correct across 5 blocks (or 20 trials); the dotted lines represents 80% correct. Data are presented separately for trials with English words as samples and trials with foreign words as samples. The English word samples during testing condition were previously comparison stimuli in the training condition, and the foreign word samples during the testing condition were previously comparison stimuli during the training condition. Figure 6 shows that, for three out of the eight subjects (S3, S6, and S8) in Group 0, the English-foreign word and foreign-English word conditional relations slowly became reversible during the course of testing. For these subjects, accuracy on the testing trials improved during the course of testing. For four of the five remaining subjects (S1, S2, S4, and S5), accuracy on the testing trials generally fluctuated around chance levels with no systematic difference in accuracy between trials involving English or foreign words as samples. S7’s performance on the testing trials was highly accurate in block 2 (trials 20-28) before deteriorating to around chance levels of accuracy.

Figure 7 presents data from the testing condition for each subject in Group 2. This figure shows that, for three out of the eight subjects (S10, S13, and S16) the English-foreign word and
foreign-English word conditional relations became reversible. For two of those subjects (S10 and S13) accuracy on the testing trials improved during the course of testing and for S16 performance was highly accurate from the beginning of the testing condition. The remaining five subjects’ performances (S9, S11, S12, S14, and S15) are more difficult to characterize briefly. S9 had highly accurate responses to English word samples, but responded around chance with foreign word samples. Performances of S11 and S12 fluctuated around chance levels of accuracy throughout the testing condition. For S15, performance was better than 75% correct for 11 out of 40 trials in block two.

Figure 8 presents data from the testing condition for each subject in Group 8. This figure shows that performance on the testing trials was highly accurate for six of the eight subjects who progressed to the testing condition. For these six subjects (S17, S18, S20, S21, S22 and S23), performance on trials involving foreign words as samples was either accurate throughout testing or more accurate later during the testing condition than trials involving English words as samples (with the exception of block 1 for S21). In addition, for four of the six subjects (S17, S18, S20, and S22), performance on trials involving foreign word samples was at or near 100% accurate from the beginning and throughout testing condition. For the two subjects who failed to show reversible conditional discriminations (S19 and S26), fluctuated around chance levels of accuracy during testing.

Figure 9 presents average latencies to respond to the sample stimulus and to select comparison stimuli for each of the three groups during testing. Average latencies are presented separately for trials involving English and foreign words. Mean latencies for Group 0, Group 2, and Group 8 are presented in the top, middle, and bottom panels, respectively. The data presented in the top and middle panels show that, for subjects in Group 0 and Group 2 there was no systematic difference in the latencies to respond to English or foreign words samples or to select from an array of comparison stimuli during testing. The data from Group 8 (bottom panel) show that these subjects took more time to select foreign word comparisons relative to
English word comparisons during testing. The bottom panel also shows that latency to respond was similar for the English word vs. the foreign word samples during testing. Average latencies to respond in the presence of all stimuli were greater for Group 8 than for Groups 0 and 2.
DISCUSSION

The purpose of the present study was to evaluate the effects of delay to comparison onset on the acquisition of conditional discriminations using a delayed match-to-sample (DMTS) procedure and the reversibility of those conditional discriminations. According to Schroth (1995) increased task difficulty improves knowledge “transfer” even as it retards acquisition. In this experiment, participants in the group that experienced longer delays consistently responded at a higher overall accuracy during testing, with 6 out of 8 subjects responding at greater than 80% accuracy, compared to 3 and 4 out of 8 participants in Group 0 and Group 2, respectively. Overall, this pattern of results suggests that increasing the delay between sample offset and comparison onset increased the likelihood of transfer of learning in the DMTS task.

Wright et al. (1988) suggested that the positive effect on transfer of training with multiple exemplars documented in their study might have been due to blocking of certain learning strategies that do not produce transfer. Specifically, Wright et al. (1988) suggested that their use of many exemplars and trial unique training may have blocked “configural” learning – a pattern of responding that has been shown to hinder the transfer of learned relations to novel stimuli (see Wright, 1997). It is possible that, in the current study, that increased delays between sample offset and comparison onset may have altered the strategies subjects employed when learning the task. By disrupting these strategies, subjects may have been able to transfer the learned conditional relations to a novel situation, a symmetry test.

Horne and Lowe (1996) suggest that naming of arbitrary stimuli is sufficient and, indeed, may be the only means by which stimulus classes are established. The most common way verbally able humans name is termed common naming, however, Horne and Lowe also describe more complex ways to name such as intraverbal naming, autoclitic naming, and rule-governed behavior (1996). Several subjects (S1, S4, S5, S8, S11, S18, S20, S21) reported using an intraverbal naming response to learn the task: Dumb house (DUUM-HOUSE), Bedpost (BED-POSTEL), Merging bank (MARGEM-BANK). With three exceptions, the subjects who
reported intraverbal naming or the use of a mnemonic device did not respond symmetrically in the testing condition. Whether this supports the naming account is unclear however since the use of intraverbal naming was not directly manipulated and not all subjects reported the use of intraverbal naming during debriefing.

The data described in our experiment may be explained in terms of Stimulus Control Topography Coherence (hereafter, SCT) Theory (Ray & Sidman, 1970; McIlvane & Dube, 2003). According to McIlvane & Dube (2003), one essential factor for SCT coherence is that the properties of the relation established by the experimenter to control subject responding must be the same as those properties that actually control the behavior of the subject. If those properties established by the experimenter do not control subject behavior or are only part of the properties that control subject behavior the subject could fail to learn the task or fail to show emergent behavior. SCT Theory may provide the most parsimonious explanation of why some subjects learned the training material and others did not. The properties of the 8 s condition established by the experimenter appear to have established control over subject responding whereby most of the subjects showed emergent symmetrical responding.

One factor that may have influenced the topography of stimulus control is the duration of contact with the sample stimuli. Average latencies to respond to the sample stimuli were longer for subjects in the 8 s condition than for those in either the 0 s or 2 s group, with the latencies to responses to the foreign sample being the longest. Identical outcomes were obtained for latencies to respond to comparison stimuli. These outcomes are consistent with the results of a study by Wright (1997), who showed that increasing the time spent with the sample stimulus could enhance conditional discrimination learning. If increased time observing stimuli improves control over responding by characteristics of the relation designed by the experimenter to exert control (e.g., formal properties of the relation), then SCT would predict more accurate responding and increased probability of emergent responding for subjects in the 8s group.
It could be that individual differences together with fortuitous assignments account for the differences observed across groups. There are two ways to address this possibility. The first is to replicate the current set of procedures with a much larger number of subjects such that statistical tests of sufficient power can be used. A second way would be to attempt to study the influence of delay in a single-subject design. Future research should pursue both avenues to increase our understanding of this phenomenon. Other research should be conducted manipulating delay values within subject and different variables related to the training of the task such as, number of exemplars, criterion to move to testing, sample requirement, etc. in order to determine what aspects of task difficulty enhance or retard transfer.

Previous research has indicated that alterations of tasks that increase the putative difficulty of tasks both disrupt acquisition and promote transfer (Schroth, 1995; Bjork & Allen 1970). Documenting the occurrence, nature, and degree of degradation in learning during acquisition was more complicated than the disruptions reported by Schroth (1995, 1997) and Wright et al. (1988). For example, a comparison of the data presented in Figures 2 - 4 shows that the mean number of trials required to meet the acquisition criterion across groups actually decreased as duration between delay to comparison onset increased. On the other hand, the proportion of subjects who met the acquisition criterion decreased as the delay between sample offset and comparison onset increased. The first set of findings is not consistent with the notion that decrements in acquisition were produced as a function of an increase in delays; however, the second set of outcomes is consistent with such an account. It may be that the delays between sample and comparison stimuli affect acquisition differently from previous manipulations (e.g., number of elements in the task [Bjork & Allen, 1970], type of training [Schroth, 1997] ) or that different results are obtained with different types of tasks (e.g., DMTS in the current experiment versus recall tasks [Bjork & Allen, 1970]). Future research may elucidate the effects, if any, of such variables on acquisition and transfer of relations among stimuli.
Table 1
Word Pairs Used During Baseline and Training

<table>
<thead>
<tr>
<th>Sample</th>
<th>Comparison</th>
<th>Sample</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>Potrava</td>
<td>Postel</td>
<td>Bed</td>
</tr>
<tr>
<td>Bathroom</td>
<td>Koupelna</td>
<td>Duum</td>
<td>House</td>
</tr>
<tr>
<td>Ball</td>
<td>Esfera</td>
<td>Comboio</td>
<td>Train</td>
</tr>
<tr>
<td>Farm</td>
<td>Fazenda</td>
<td>Margem</td>
<td>Bank</td>
</tr>
</tbody>
</table>
Table 2

Word Pairs Used During Testing

<table>
<thead>
<tr>
<th>Sample</th>
<th>Comparison</th>
<th>Sample</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potrava</td>
<td>Food</td>
<td>Bed</td>
<td>Postel</td>
</tr>
<tr>
<td>Koupelna</td>
<td>Bathroom</td>
<td>House</td>
<td>Duum</td>
</tr>
<tr>
<td>Fazenda</td>
<td>Farm</td>
<td>Bank</td>
<td>Margem</td>
</tr>
<tr>
<td>Esfera</td>
<td>Ball</td>
<td>Train</td>
<td>Comboio</td>
</tr>
</tbody>
</table>
FIGURE CAPTIONS

Figure 1. Percent correct in baseline probes for English and foreign word samples for each subject for Groups 0, 2, and 8. Asterisk denotes 0% for those trials.

Figure 2. Percent correct across blocks during training condition for Group 0.

Figure 3. Percent correct across blocks during training condition for Group 2.

Figure 4. Percent correct across blocks during training condition for Group 8.

Figure 5. Average latency in seconds for sample and comparison stimuli during the training condition for each group. Sample and comparison latencies are further divided into English and foreign word stimuli.

Figure 6. Percent correct across blocks during testing condition for Group 0. The dotted line represents 80%.

Figure 7. Percent correct across blocks during testing condition for Group 2. The dotted line represents 80%.

Figure 8. Percent correct across blocks during testing condition for Group 8. The dotted line represents 80%.

Figure 9. Average latency in seconds for sample and comparison stimuli during the testing condition for each group. Sample and comparison latencies are further divided into English and foreign word stimuli.
Figure 1.
Figure 2.
Figure 3.
Figure 4.
Figure 5.
Figure 6.
Figure 7
Figure 8.
Figure 9.
REFERENCES


Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination

