Published in final edited form as:

J Appl Behav Anal. 2017 January ; 50(1): 87–105. doi:10.1002/jaba.363.

Reducing Overselective Stimulus Control with Differential Observing Responses

Rachel S. Farber,

E. K. Shriver Center, University of Massachusetts Medical School; Department of Psychology, Western New England University.

Chata A. Dickson, and

New England Center for Children; Department of Psychology, Western New England University.

William V. Dube

E. K. Shriver Center, University of Massachusetts Medical School; Department of Psychology, Western New England University.

Abstract

Overselective stimulus control refers to discriminative control in which the number of controlling stimuli is too limited for effective behavior. Experiment 1 included 22 special-education students who exhibited overselective stimulus control on a two-sample delayed matching task. An intervention added a compound identity matching opportunity within the sample observation period of the matching trials. The compound matching functioned as a differential observing response (DOR) in that high accuracy verified observation and discrimination of both sample stimuli. Nineteen participants learned to perform the DOR and two-sample delayed matching accuracy increased substantially for 16 of them. When the DOR was completely withdrawn after 10 sessions, accuracy declined. In Experiment 2, a more gradual withdrawal of DOR requirements showed that highly accurate performance could be maintained with the DOR on only a proportion of trials for most participants. The results show that DOR training may lead to a general improvement in observing behavior.

Keywords

differential observing responses; intellectual disabilities; matching to sample; overselective stimulus control

Overselective stimulus control (also known as stimulus overselectivity; Lovaas, Schreibman, Koegel, & Rhem, 1971) refers to discriminative stimulus control in which the number of controlling stimuli or stimulus features is too limited for effective behavior (for reviews, see Lovaas, Koegel, & Schreibman, 1979; Ploog, 2010). Children with intellectual disabilities may exhibit overselective stimulus control that interferes with learning academic and communication skills (reviewed in Dube, 2009). For example, if stimulus control is

restricted to the initial letters in words, then the student may make errors in discrimination of words with letters in common like CAT, CAR, CAP (e.g., Walpole, Roscoe, & Dube, 2007; cf. Yoo & Saunders, 2014). Similarly, students learning to use augmentative and alternative communication methods such as the Picture Exchange Communication System (Bondy & Frost, 2001) may make errors if stimulus control is restricted to single features within symbols, or to single symbols within multi-symbol phrases (Dube & Wilkinson, 2014).

Overselective stimulus control in matching-to-sample tasks that are widely used in discrete-trial instruction may result in intermediate accuracy scores, high enough above chance level to indicate stimulus control by some samples or features of samples, but low enough to indicate that not all samples or features exert control (e.g., 67% correct on a three-choice matching task; Dickson, Deutsch, Wang, & Dube, 2006). The present study addressed overselective stimulus control by sample stimuli as it may occur in matching to sample with two sample stimuli per trial. In typical matching procedures, the sample stimuli are presented first, the student makes an observing response to the samples, and then an array of comparison stimuli is presented, one of which is correct in relation to the samples. Observing responses to sample stimuli may be classified as nondifferential or differential. With nondifferential observing responses, the response to samples is the same on every trial and the only discrimination required is between the presence versus absence of the stimuli. By contrast, differential observing responses (DORs) include behavioral requirements that verify discrimination of stimulus features that differ among the samples. For example, if the sample stimuli are pictures, one type of DOR would be to name the pictures aloud (e.g., Constantine & Sidman, 1975). The intervention described in the present study involved modifying the requirements for observing responses to the sample stimuli.

Nondifferential observing responses are not effective for reducing overselective stimulus control with multiple sample stimuli (e.g., Reed, Altweck, Broomfield, Simpson, & McHugh, 2012). By contrast, the inclusion of DORs has been shown to greatly reduce overselective stimulus control in matching tasks with multiple samples. For example, in a two-sample matching task with pictures, a requirement that participants name each sample stimulus aloud before proceeding with the trial has been effective (e.g., Gutowski & Stromer, 2003). Naming the sample stimuli, however, requires that the participants have previously learned the names, and thus would not be useful in situations with novel stimuli or for learners who are unable to produce names. Another type of DOR, compound identity matching, may be more broadly useful because it requires only that participants are able to perform generalized identity matching (in which the correct comparison stimulus is physically identical to the sample). For example, if the sample stimuli for a two-sample matching trial were A+B, the comparisons for a compound identity-matching trial might be A+B (correct), A+C, and D+B (both incorrect). Because each of the incorrect comparisons has one stimulus in common with the sample, (a) overselective stimulus control by only one of the samples would lead to errors on some trials and thus (b) consistently high accuracy over a series of trials verifies stimulus control by both of the stimuli in the sample compound.

In previous research, Dube and McIlvane (1999) showed that this type of compound matching task could be used as a DOR to reduce overselective stimulus control. Participants

were three special-education students with intellectual disabilities. The baseline condition was delayed matching to sample with two samples on every trial and a nondifferential observing response, a single touch to the sample stimulus display. The stimuli were abstract forms created in the research laboratory. Participants exhibited overselective stimulus control, with mean accuracy scores in the 65-79% range. For the DOR intervention, the observing response procedure was elaborated to include simultaneous compound identity matching as described above. In the first intervention session, accuracy immediately increased to 86-91% and these improvements maintained for 10 sessions. The goals of Experiment 1 in the present study were to assess the compound-matching DOR procedure's general utility by replicating Dube and McIlvane with a larger number of participants and to extend research with the compound-matching DOR to include educationally relevant stimuli.

Dube and McIlvane (1999) abruptly ended the DOR intervention after 10 sessions and the baseline nondifferential observing response procedure was reinstated. Accuracy declined to pre-intervention levels for all three participants. The goal of Experiment 2 in the present study was to determine whether a more lasting benefit could be obtained if the DOR requirements were gradually withdrawn. Even if DOR requirements could not be completely eliminated, high overall accuracy with the DOR on only some proportion of trials would be a step toward a general improvement in observing behavior.

Experiment 1

Participants were pretested and selected for this study if they had (a) intermediate accuracy indicative of overselective stimulus control on a delayed matching-to-sample task with compound sample and single-comparison stimuli, and (b) high accuracy on a simultaneous-matching task with compound sample and comparison stimuli. The experiment began with a baseline condition of delayed matching with compound samples and single comparisons, followed by a DOR intervention condition in which the simultaneous compound-matching task was embedded within the sample observation portion of each trial, and then a return to the baseline (A-B-A reversal). Some participants received a second intervention and return to baseline (A-B-A-B-A). If the DOR intervention was effective, then the participant continued to Experiment 2 in which the intervention was reinstated.

Method

Participants—Participants were 22 special-education students with intellectual disabilities who were selected for this study on the basis of pretest results (details below). Table 1 shows participant characteristics. The Peabody Picture Vocabulary Test 4 (Dunn, Dunn, & Pearson Assessments, 2007) is a measure of receptive vocabulary. The Differential Abilities Scales II (DAS-II; Elliott, 2007), General Conceptual Ability is a standardized score similar to an IQ score ($M = 100$, $SD = 15$) that is used here to describe level of intellectual disability.

General procedures

Setting and apparatus: Testing was conducted in quiet areas containing a table and two chairs at the participant's school. Each participant sat at the table in front of a 15-in (38.1-cm) LCD color touch screen monitor (Elo 1515L); the experimenter sat behind and to one

side of the participant. Matching-to-sample trials were presented and responses (touches) recorded by a laptop computer running custom software.

Stimuli: Two sets of stimuli were used. The *color symbols* set consisted of 180 Mayer-Johnson Picture Communication Symbols (Mayer-Johnson, 2008), approximately 1.5×1.5 cm. The *arbitrary forms* set consisted of 180 black forms approximately 1.5×1.5 cm constructed for experimental purposes and presumed to be unfamiliar to participants (e.g., Dube & McIlvane, 1999; Johnson, Meleshkevich & Dube, 2014). Each session included stimuli from only one of these sets; they were never intermixed within sessions. All sessions were conducted with trial-unique stimulus presentation. The stimuli for each session were drawn at random without replacement from the pool of 180 stimuli, and different stimuli were presented on every trial.

Matching to sample procedure and trial type designations: The procedure was three-choice identity matching to sample. Trials began when a sample was presented in the center of the screen. There was no limit for the sample presentation duration; the sample remained displayed until the participant touched it. When the participant touched the sample: (a) in simultaneous matching to sample, the sample remained displayed and three comparison stimuli were presented in a horizontal array at the top of the screen; (b) in delayed matching to sample, the sample disappeared and three comparison stimuli were presented immediately at the bottom of the screen (i.e., 0-s delay). The correct comparison was identical to a sample and neither of the two incorrect comparison stimuli was identical. The correct response was touching the identical comparison. Within each session, the order of correct comparison locations was unsystematic with the restrictions that a specific location was repeated on no more than three consecutive trials and each comparison position was correct equally often. The overall duration of test sessions was approximately 10-15 min. One to three sessions were conducted on the same day, depending on participants' availability and scheduling constraints.

The experiment included several variations of matching-to-sample procedures, both simultaneous and delayed, and with samples and comparisons consisting of one or two stimuli. For clarity and convenience, we will refer to the trial types with a modification of the trial designations of Cox and D'Amato (1982; also Stromer, McIlvane, Dube, & Mackay, 1993). An initial letter S or D will designate simultaneous or delayed matching, respectively. Second and third letters S or C will designate single (one stimulus) or compound (two stimuli) samples and comparisons, respectively. For example, D-CS refers to delayed matching with compound samples and single comparisons; see Figure 1 for further examples.

Consequences: During all sessions, every correct response was followed by a 2-s on-screen display of animated stars, computer-generated chimes, and presentation of a poker-chip token by the experimenter. After the stars display, the screen was blank during a 3-s inter-trial interval, and then the next trial began with a sample presentation. Incorrect responses were followed by a 2-s black screen and then the inter-trial interval.

Prior to participation, caregivers were asked to suggest preferred items used in the classroom for each individual participant, including snack foods, money, trinkets, and activities (e.g., playing a computer game). Accumulated tokens were exchanged for the participant's choice(s) among the available items after each session. The price of each preferred item was 10 tokens and this remained the same across sessions. Before each session, the preferred items were displayed on a countertop for the participant to see; in front of each item was a clear plastic tube that held 10 tokens. Following each session, the participant selected an item and filled the corresponding tube to the top; the experimenter collected the tokens from the tube and presented the item to the participant. If there were tokens remaining after filling the first tube, the participant chose another item and filled the corresponding tube. This process continued until the participant exchanged all of the accumulated tokens. If the participant earned an amount not divisible by 10, the price of the last item would be adjusted for that session. For example if a participant earned 36 tokens, three items would be priced at 10 tokens and the last item would be adjusted to six, using up all of the tokens the participant earned.

Pretests

One-sample pretest: D-SS: This pretest verified accurate delayed identity matching with single sample and comparison stimuli. Sessions consisted of 18 trials with the color symbols stimulus set. The criterion to complete this pretest was one session with at least 89% accuracy (16/18 correct). Sessions were discontinued if this criterion was not met within three sessions and there was no increasing trend in accuracy (this occurred with five potential participants, data not reported here).

Two-sample pretest: D-CS: This pretest assessed possible overselective stimulus control when there were two sample stimuli on each trial. Sessions consisted of 36 trials with the color symbols stimulus set. The two sample stimuli were presented approximately 1 cm apart. Only one of the two sample stimuli appeared in the comparison array. For example, if the sample stimuli were AB, the comparison array might be {A, X, and Y} or {B, X, and Y}. The left and right sample stimulus was correct equally often. Each participant experienced three sessions.

The criterion to advance to an experimental condition was a 3-session mean accuracy score within an overselective range for the task, defined as accuracy between 50% and 83%. The rationale for this range is: Overselective stimulus control by only one sample per trial would result in (a) approximately 100% correct responding when that stimulus appeared as the correct comparison, but (b) chance level of 33% when that stimulus did not appear in the comparison array. With both types of trials equally represented, the overall accuracy score would be approximately 67% (see Dickson et al., 2006 for further details of the accuracy score analysis). The overselective range lower limit of 50% is halfway between 67% and chance (33%), and the upper limit of 83% is halfway between 67% and 100%.

If a participant's 3-session mean accuracy score was less than 50%, participation was discontinued (one potential participant, data not reported here). If mean accuracy was greater than 83% and thus not within the overselective range with the color symbols, the D-SS and

D-CS pretests were repeated with the arbitrary forms stimulus set. Previous research has shown that D-CS accuracy scores are likely to be lower with the arbitrary forms stimulus set than with the color symbols (Dube et al., 2016). If mean accuracy was greater than 83% with the arbitrary forms, participation was discontinued (six potential participants, data not reported here).

Compound matching pretest: S-CC: This pretest assessed whether participants could perform compound matching with high accuracy in a simultaneous identity-matching task that allowed them to look back and forth among sample and comparison stimuli. Sessions consisted of 36 trials with the same stimulus set that yielded D-CS accuracy in the overselective range. The number, size, and distance between sample stimuli on these trials were the same as those for the D-CS pretest. Each of the three comparisons also consisted of two stimuli. The correct comparison was identical to the sample, and each of the two incorrect comparisons had one stimulus in common with the sample and one that was not part of the sample. One of the incorrect comparisons shared a common stimulus with the left sample stimulus and the other incorrect comparison shared a common stimulus with the right sample stimulus. For example, if the sample stimuli were AB, the comparison array might be {AB, XB, and AY}. The criterion to advance to the experimental conditions was three consecutive sessions with at least 89% accuracy (32/36 correct). Sessions were discontinued if this criterion was not met within 12 sessions and there was no increasing trend in accuracy.

Experimental conditions

D-CS baselines: Sessions consisted of 36 trials with the same procedure as the D-CS pretest. The initial baseline condition (Baseline 1) was arranged as a multiple-baseline design across participants with 5, 8, or 11 baseline sessions (see Table 2). If there was an increasing trend in accuracy scores, additional sessions were conducted until accuracy was stable. After the DOR intervention condition(s), the return to baseline condition(s) (Baseline 2 and for some participants Baseline 3) was a minimum of 10 sessions.

DOR intervention: Sessions consisted of 42 trials (36 DOR and 6 S-CC). For DOR trials the S-CC task was embedded within the sample observation portion of the D-CS trial. The first touch to the compound sample produced the S-CC comparison array at the top of the screen. After a touch to any of the compound comparisons, the array disappeared. There were no differential consequences for the S-CC portion of the trial so that there was no programmed reinforcement for selecting comparisons before the D-CS portion of the trial. After a second touch to the sample, the sample stimuli disappeared and the D-CS comparison array was presented at the bottom of the screen. So that participants could not select the correct comparison simply by excluding novel stimuli, the incorrect comparisons for D-CS portion of the trial were the same stimuli that had appeared within the incorrect compound comparisons in the S-CC portion of the trial. For example, if the sample stimuli were AB and the comparisons in the S-CC portion of the trial were {XB, AY, and AB}, then the comparisons in the D-CS portion of the trial were {A, X, and Y} or {B, X, and Y}.

To maintain accurate performance on the S-CC portion of the trials, six trials in each DOR session were S-CC trials that ended after the response to the comparisons with differential consequences for correct and incorrect responses. These 6 trials were interspersed randomly throughout the 36 DOR trials.

The goal of the DOR intervention was to evaluate functional use of the S-CC task as a DOR. This intervention depended on maintaining high accuracy on the S-CC responses. If accuracy were to decline, the errors would indicate that the participant was no longer consistently emitting DORs. To monitor this, accuracy scores for each session were calculated separately for the 42 S-CC responses (36 as part of DOR trials plus 6 S-CC trials) and the 36 D-CS responses. If accuracy on the S-CC responses was below 89% in any session, the following session was a remedial session to recover S-CC performance. Remedial sessions included 36 S-CC trials only, with differential consequences (as in the S-CC pretest). The criterion to return to DOR sessions was one remedial session with at least 89% accuracy. If this criterion had not been met within three consecutive remedial sessions, participation would have been discontinued; this did not occur. The DOR condition continued for 10 sessions with at least 89% accuracy on S-CC responses, or to a limit of 6 DOR sessions in which S-CC accuracy fell below 89%.

Results and Discussion

Table 2 shows stimulus sets, number of Baseline-1 and remedial sessions, and mean accuracy scores for all 22 participants. Because participants received different numbers of sessions in Baseline 1, the accuracy scores were calculated for the last five sessions (range, 61-82%). Three participants (A31, B15, and B78) did not maintain S-CC accuracy of at least 89% during six DOR sessions. After each such session, the participant received one or more remedial sessions. After six DOR sessions in which S-CC accuracy was less than 89%, these participants were withdrawn from the study and did not complete the DOR intervention condition. This result shows that accurate performance in sessions that include only S-CC trials with differential consequences on every trial may not be sufficient to establish the behavioral prerequisites for using the S-CC task as a DOR. Additional training may be necessary in some cases. The failures to maintain S-CC accuracy were unrelated to participant's PPVT and DAS scores. Table 1 shows that the scores for Participants A31, B15, and B78 were not near the lower end of the distributions for these scores, and thus the results of these standardized tests may be poor predictors of compound matching with the S-CC task.

The other 19 participants completed 10 DOR intervention sessions with S-CC accuracy of at least 89% (Table 2). For 8 of these 19 participants, accuracy on S-CC responses in DOR sessions fell below 89% in at least one session, and these sessions were followed by a remedial session; accuracy always recovered to at least 89% after one remedial session.

The D-CS accuracy scores during DOR intervention are the means for those 10 sessions with S-CC accuracy of at least 89%. The median D-CS accuracy score for all participants increased from 72% in Baseline 1 to 91% in DOR intervention. For 16 of these 19 participants, the intervention was effective in that mean D-CS accuracy increased to levels above the overselective range (> 83%; exceptions were Participants B65, B91, and C15). In

addition, nine participants also met a more stringent mastery criterion of three consecutive sessions with accuracy at least 89% within the last five sessions of the condition. Figure 2 shows individual session data in baseline and intervention conditions. Figure 2A shows data for three representative participants (B06, B82, and B18) selected for this figure on the basis of mean Baseline 1 and DOR intervention accuracy scores that approximated the medians for all participants.

Figure 2B shows individual session data for Participant B65, representative of the three participants whose accuracy during intervention did not improve to a level above the overselective range (the others were B91 and C15). For these participants, the high S-CC scores indicate that they were making DOR responses, but the D-CS scores indicate that it was not an effective intervention.

The Baseline 2 results in Table 2 and Figure 2A show that the 10-session exposure to the DOR intervention produced little net improvement in most cases. Median D-CS accuracy for Baseline 2 was 77%, a very modest increase over the Baseline-1 median of 72%. Figure 2C shows individual session data for one exception, Participant B84. Accuracy was variable in Baseline 2 and most scores were above the overselective range. Because of the variability, additional sessions were conducted and B84 met the mastery criterion with no DOR procedure.

A second DOR intervention condition was given to a subset of seven participants who did not meet the mastery criterion during the first intervention and who were available for further testing. Table 3 shows the results; the Baseline 2 data are repeated from Table 1. As before, D-CS accuracy improved during DOR intervention when the DOR was included in every trial. For three participants, S-CC accuracy fell below 89% in one or more DOR sessions, but accuracy always recovered to at least 89% after one remedial session. During DOR intervention, median D-CS accuracy was 91%, six of the seven scores were above the overselective range, and five participants met the mastery criterion for the first time. As with the first DOR intervention, however, the improved accuracy did not persist when the intervention was withdrawn; the median accuracy score for Baseline 3 was 78%. Figure 2D shows a representative example of individual session results.

To summarize, the DOR intervention was implemented with 19 of 22 participants and it was effective for eliminating overselective stimulus control with 16 of these 19 participants. A total of 14 participants met the mastery criterion during the first or second intervention. These results indicate that S-CC matching may be a generally useful intervention for overselective stimulus control in two-sample matching for those students who can maintain accurate compound matching. The results replicate and extend those of Dube and McIlvane (1999) to include a larger number of participants and to demonstrate successful interventions for overselective stimulus control with two-stimulus sample arrays of Mayer-Johnson symbols used in augmentative and alternative communication.

Experiment 2

In Experiment 1, the intervention condition included the DOR requirement on every D-CS trial. Although this was effective in the majority of cases, it also doubled the number of responses per session and thus substantially increased both the response effort and the time required. For some participants, such a high level of instructional support may not be necessary; occasional exposures to the DOR may be sufficient and more efficient.

One interpretation of the results is that the DOR procedure functioned as a prompt to observe both of the sample stimuli. If so, the decrease in accuracy when the DOR was discontinued may be analogous to a failure of stimulus control transfer from a prompt to a target. Transfer from prompt to target may sometimes be facilitated by gradually withdrawing the prompt. In Experiment 2, we explored this approach with a titration procedure that adjusted the proportion of DOR trials within sessions, decreasing that proportion if accuracy was high, and increasing it if accuracy declined. The goal was (a) to determine whether the DOR could be completely eliminated while maintaining high accuracy on the D-CS task, or (b) if the DOR could not be completely eliminated, to conduct a parametric analysis to determine a threshold level of DOR support that would maintain high accuracy.

The titration procedure was implemented as a prompt withdrawal program with six levels of DOR support. These titration levels were presented in relatively brief blocks of 12 or 14 trials to allow frequent and within-session adjustments in level of support as needed. Each session included two or three blocks of trials. A software algorithmic steering logic was used to evaluate performance after each block of trials and decrease to a lower level of support if accuracy was high, continue with the current level if accuracy was intermediate, or increase to the previous level if accuracy declined. The software was thoroughly tested and debugged before the experiment to ensure that procedural adjustments were consistent with the criteria detailed in the Methods section below. The titration process continued until stability criteria were met. Following the DOR titration condition, a posttest was conducted to evaluate whether the level of DOR support estimated by the titration procedure was sufficient to maintain high D-CS accuracy under conditions like those of Experiment 1.

Method

Participants—Participants were 11 individuals who completed Experiment 1. They were selected for this experiment if (a) they had met the mastery criterion of three consecutive sessions with D-CS accuracy of at least 89% while the DOR intervention was in effect, but (b) accuracy declined to the overselective range when the DOR was discontinued. See Table 4 for these participants' identification codes and Table 1 for participant characteristics.

Procedures—Setting, apparatus, stimuli, and consequences were as described in Experiment 1. Matching-to-sample procedures were also as described in Experiment 1 (depicted in Figure 1): D-CS, S-CC with differential consequences, and the DOR procedure with the S-CC task embedded within the sample observation portion of a D-CS trial.

DOR titration: Each session was divided into two or three blocks of 12 or 14 trials. After each block, the software's algorithmic steering logic evaluated the results, prepared the next block of trials, and prompted the experimenter to press a key indicating whether to continue or end the session. If the session ended, the information for the next block of trials was saved and the following session began with that prepared trial block.

The titration trial blocks were structured as a series with six levels of DOR support, shown in Table 5. The proportion of DOR trials versus D-CS baseline trials (no DOR) per block decreased as the level number decreased, in a manner analogous to fading out a prompt. Level 5 was similar to the DOR intervention in Experiment 1, with 12 DOR trials and no D-CS baseline trials. Level 4 included nine DOR trials and three D-CS baseline trials, Level 3 included six of each trial type, and Level 2 three DOR trials and nine D-CS baseline trials. In Levels 4-2, DOR and D-CS baseline trials were presented in random order. In Levels 1 and 0 there were no DOR trials.

Levels 5-1 also included two S-CC trials with differential consequences. As in Experiment 1, these trials were added to help maintain the S-CC performance, and also to provide occasional immediate reinforcement for observing both sample stimuli. One of these trials was always the first trial in the block, and the second was presented in random order. There were no S-CC trials in Level 0.

After each block of titration trials, the software steering logic separately evaluated S-CC and D-CS responses and prepared the next block of trials according to the criteria described below. S-CC responses within each block of trials included the two S-CC trials with differential consequences (Levels 5-1) and the S-CC portions of any DOR trials (Levels 5-2). If a participant made two or more S-CC errors, the subsequent block of trials was a remedial block consisting of 12 S-CC trials with differential consequences. The criterion to return to the previous titration level was one remedial block with at least 11/12 correct. If this criterion had not been met within three consecutive remedial blocks, participation would have been discontinued; this did not occur.

The D-CS responses evaluated after each block of trials included both the final responses on any DOR trials (Levels 5-2) and the responses on any D-CS baseline trials (Levels 4-0), a total of 12 responses per block. The program progressed to the next lower level following three consecutive blocks with at least 11/12 correct; continued with the current level following each block with 10/12 correct (with one exception); and backed up to the previous level following either (a) one block with less than 10/12 correct or (b) six blocks with 10/12 correct within a series of blocks on the same level (the exception). The DOR titration condition terminated after either (a) six back-ups at the same level or (b) six consecutive blocks with at least 11/12 correct on Level 0.

We used the titration results to estimate the threshold level of DOR support necessary for accurate D-CS in longer sessions like those of Experiment 1 (36-42 trials). An overall mean accuracy score was calculated for each titration level, based on all D-CS responses throughout the entire series of titration sessions; this included both DOR and D-CS baseline trials, but did not include responses in any blocks with two or more S-CC errors. The

threshold was estimated to be between the lowest level with an overall mean accuracy of 89% or greater and the next lower level. For example, if accuracy was high at Level 3 (with 6 DOR trials per block) but fell below 89% at Level 2 (with only 3 DOR trials per block), then the threshold was between Levels 3 and 2. Participants who completed titration with six consecutive blocks on Level 0 with at least 11/12 correct were exceptions in that there were no DOR trials and thus no threshold.

Posttest: We conducted a posttest to evaluate the accuracy of the threshold estimates from the DOR titration condition. Posttest sessions consisted of 36 or 42 trials, as in Experiment 1, and the mixture of trial types within sessions was organized as six levels, as in the titration procedure. The proportions of different trial types were the same as those for the titration levels and as shown in Table 5, except there was three times as many of each trial type per session. For example, in posttest Level 4, sessions included 27 DOR trials, 9 D-CS baseline trials, and 6 S-CC trials with differential consequences, for a total of 42 trials. Thus, posttest Level 5 sessions were identical to the DOR intervention sessions in Experiment 1 with DOR responses on every trial, and posttest Level 0 sessions were identical to D-CS baseline sessions in Experiment 1 with no DOR responses on any trials. Posttest Levels 4-1 consisted of intermediate proportions of trials with DOR requirements. If the estimated threshold was correct, then accuracy on posttest sessions would be at least 89% on the posttest level above the threshold and less than 89% on the posttest level below the threshold.

Posttest sessions continued until a posttest threshold was identified using criteria similar to those for the titration procedure: (a) increasing the level if D-CS accuracy was below 83% for two consecutive sessions, (b) decreasing the level if accuracy was 89% or greater for three consecutive sessions, and (c) increasing the level if neither of these criteria were met within 10 sessions. As in Experiment 1, S-CC remedial session(s) were conducted if accuracy on S-CC responses fell below 89% during a posttest session.

Results and Discussion

DOR Titration—Table 4 shows the results of the DOR titration procedure for each participant. The total number of titration blocks ranged from 75 to 138. The table also shows the number of S-CC remedial blocks each participant received; these blocks followed titration blocks with two or more S-CC errors. S-CC matching was generally very reliable. Four participants never required S-CC remedial blocks, and no participant required more than four remedial blocks.

The closed circles in Figures 3 and 4 show mean D-CS accuracy for each level across the entire DOR titration condition. The dotted line in each plot shows the accuracy criterion of 89%. Data points at or above the dotted line show the levels for which the DOR support was sufficient to maintain high D-CS accuracy. Estimated DOR thresholds are shown by adjacent points that cross this line as the level decreases. For example, A34's accuracy was 92% for Level 3 (with DOR on 50% of the trials) but 83% for Level 2 (with DOR on 25% of the trials), and thus the estimated threshold was between Levels 3 and 2. For Participant A93 there were two instances of adjacent points that crossed the accuracy criterion line as the

level decreased (Levels 5-4 and Levels 3-2) and the threshold estimate was based on the lower of these (Levels 3-2).

Posttest—The open circles in Figures 3 and 4 show mean accuracy for the last three posttest sessions for each posttest level that was tested. Figure 3 includes posttest results for seven participants for whom the required level of DOR support was accurately estimated by the titration procedure. For example, the titration results for A34 indicated a threshold between Levels 3 and 2. This was confirmed by posttest accuracy of 89% on Level 3 and 86% on Level 2. The figure also shows that the level of DOR support needed to maintain high D-CS accuracy varied across individuals.

There was no threshold estimate for Participants B68 and B75 because they met the accuracy criterion for titration Level 0 (with no DOR or S-CC trials). This titration result was confirmed in posttests by high accuracy scores on posttest Level 0, in sessions that were identical to those of the baseline conditions in Experiment 1. For these two participants, experience with the gradual withdrawal of DOR requirements during the titration procedure apparently eliminated overselective stimulus control on the D-CS task, with increases in accuracy from 80% in Baseline 3 to 95% in posttest for B68, and from 70% to 93% for B75.

The top panel of Figure 4 shows the posttest results for Participant A93, for whom the required level of DOR support was underestimated. The titration results indicated a threshold between Levels 3 and 2, and thus that DOR requirements on 50% of the trials should be sufficient to maintain high accuracy on the D-CS task. Posttest results, however, showed that A93 met the accuracy criterion only on Level 4, with DOR requirements on 75% of the trials. The bottom three panels in Figure 4 show posttest results for three participants for whom the required level of DOR support was overestimated (B06, B66, and B89). For example, titration results estimated that B06 would require the DOR intervention on every trial (Level 5) to maintain high D-CS accuracy. Posttest results, however, showed that B06 also met the accuracy criterion on Levels 4 and 3, in which the DOR was scheduled for only 75% or 50% of the trials in the session.

To summarize in terms of the Experiment 2 goals: (a) The DOR was completely eliminated while maintaining high accuracy on the D-CS task for 2 of 11 participants. (b) The DOR could not be completely eliminated for the remaining nine participants, and the titration analysis successfully determined a threshold level of DOR support that would maintain high accuracy for five of these participants. The titration analysis underestimated the required level of DOR support for one participant and overestimated it for three participants, but sufficient levels of support were identified in the posttests for all participants.

General Discussion

In Experiment 1, 19 participants with baseline D-CS accuracy scores in the overselective range completed one or two 10-session DOR interventions. While the DOR intervention was in effect, accuracy increased above the overselective range ($> 83\%$) for 16 participants, and 14 of these participants also met a more stringent mastery criterion of three consecutive sessions with accuracy at least 89% during their first or second exposure to the DOR

intervention (e.g., Figures 2A and 2D). These results extend those of Dube and McIlvane (1999) by showing that the compound-matching DOR approach can be a generally effective intervention for two-sample delayed matching in a majority of participants with overselective stimulus control on this task. The present results also demonstrate effective DOR intervention with educationally relevant stimuli, Mayer-Johnson Picture Communication Symbols.

After the relatively brief 10-session exposures to the DOR procedure in Experiment 1, accuracy declined to pre-intervention levels when the DOR was discontinued (Participant B84 was the only exception). One interpretation of this result is related to the stimulus control of effective observing behavior on DOR trials. As with all responses to sample stimuli in baseline conditions, the first response to the sample stimuli on DOR trials was a nondifferential observing response. It was necessary for the participant to observe only that the monitor screen was no longer blank. Presentation of the comparison array on the S-CC part of the DOR trials may have been the controlling stimulus that functioned as a prompt to observe both of the sample stimuli. The DOR procedure may thus have provided reinforcement for a behavioral chain in which the initial response to the sample stimuli occurred before both of those stimuli had been observed. When the DOR requirement was discontinued, however, the initial response to the samples resulted in their disappearance and the opportunity for further observation was lost. This interpretation of stimulus control of observing behavior topographies by the DOR procedure may be tested in future research by using an eye-tracking apparatus to directly measure observing behavior (e.g., Dube et al., 2010).

In Experiment 2, the titration procedure gradually adjusted the proportion of DOR trials within sessions, decreasing that proportion if accuracy was high, and increasing it if accuracy declined. With this approach, only two participants, B09 and B18, continued to require the DOR on every trial to maintain high accuracy (see Figure 3). The DOR was completely eliminated for two participants, B68 and B75, and partially eliminated for seven others. Thus, 9 of 11 participants emitted unprompted and effective observing of two sample stimuli on at least some trials. For example, the posttest mean accuracy score for Participant B82 was 97% at Level 3 with the DOR on half of the trials in the session, but 82% on Level 2 with the DOR on only one-quarter of the trials (Figure 3). The high score for Level 3 indicates that unprompted and effective observing behavior occurred and was followed by reinforcement on approximately half of the trials (those without the DOR). One question for further research concerns the effects of continued training with reinforcement for effective observing behavior. For example, would continued training for Participant B82 at Level 3 for some extended period of time increase the behavioral persistence of unprompted and effective observing behavior? If so, then a reassessment with the titration procedure after extended training may show high accuracy maintained with fewer DOR trials per session (an outcome analogous to the long-term strengthening of functional communication behavior in Wacker et al., 2011).

One limitation in the present study is that the DOR intervention did not eliminate overselectivity for 3 of 19 participants in Experiment 1 (B65, B91, and C15; e.g., Figure 2B). Consistently high accuracy on the S-CC portion of the DOR trials indicated reliable

stimulus control by both of the sample stimuli, but there were only small or moderate improvements in accuracy for the D-CS responses at the end of the trials. One possibility is that the participant sometimes selected a stimulus at the end of the trial that was different from either of the sample stimuli because there was no consequence confirming that the observing response was correct. Another possibility is related to the concept of “separable compounds” (Stromer, McIlvane, & Serna, 1993; brief review in McIlvane, 2013). The term describes findings from stimulus equivalence research showing that discrimination training with compound stimuli can lead to emergent stimulus control by the individual elements of the compound stimuli (e.g., Debert, Matos, & McIlvane, 2007; Markham & Dougher, 1993). The intervention described in the present study is consistent with this idea; stimulus control by the two-element compound is sufficient for accurate matching on the S-CC portion of the DOR trial, but stimulus control by individual sample stimuli is required for accurate matching on the D-CS response at the end of the trial. That is, the intervention is effective if the samples function as separable compounds. The results with these three participants could be described as stimulus control of the S-CC response by the two sample stimuli as a unitary compound and not by the individual stimuli (i.e., configural stimulus control, McIlvane, 2013).

A second limitation is that three participants in Experiment 1 did not complete the DOR intervention condition because high accuracy on the S-CC portion of the trials was not maintained (Table 2, Participants A31, B15, and B78). For these three participants, every time accuracy declined, it recovered during one or more subsequent sessions consisting solely of S-CC trials with differential consequences, but then declined again after a return to the DOR procedure. Possibly the sudden shift from continuous token reinforcement for all S-CC responses in remedial sessions to token reinforcement on only 6 of 42 S-CC responses in the DOR sessions was disruptive. If so, then one training strategy may be to first establish reliable S-CC performance with continuous reinforcement and then introduce the DOR procedure by gradually increasing the proportion of DOR trials per session (and thus gradually thinning the reinforcement schedule for S-CC responses).

For applications to discrete-trial instruction, the compound identity-matching approach has an advantage in that the only prerequisite skill is generalized identity matching to sample. (An instructional approach for expanding generalized identity matching with single stimuli to compound matching is described in Farber, Dube & Dickson, 2016.) For students who can perform generalized identity matching, this type of DOR could be incorporated into a variety of tasks. For example, Dube (2009) included data from a special-education student with moderate intellectual disability who was learning to name sets of printed words. New words were introduced in the context of a cumulative baseline that also reviewed previously learned words. At one point in the instruction, the student had learned to name the words EXIT, MEN, WALK, and SUN with high accuracy. When the new word MUG was introduced, however, he began to make errors by naming MUG as “Men.” Stimulus control appeared overselective and restricted to the initial letters of the words. In such cases, a DOR could be arranged by presenting a compound-matching trial with each printed word, just prior to the naming response. For example, before naming the word MUG, the student could be given a compound matching trial with the sample MUG and comparisons MUG, MEN, and DUG.

Compound identity matching is only one of several ways to implement DORs. As noted in the introduction to this paper, other behavioral procedures can also function as DORs if they (a) control observing behavior and (b) verify discrimination of the relevant stimulus features. One possibility for future research is to conduct comparative analyses of compound matching with alternatives such as naming stimuli aloud, repeating auditory stimuli aloud (echoic), or reproducing visual stimuli by drawing them or typing the letters of words on a keyboard (Dube, 2009). Another research possibility is to compare the DOR approach to one that increases the sample stimulus display duration (and thus the opportunity for observing) by increasing the number of nondifferential observing responses on each trial. For example, Doughty and Hopkins (2011) compared 1 versus 10 nondifferential observing responses per trial on a two-sample delayed matching task with an adult with autism. The observing responses were mouse clicks over the sample. In two replications, intermediate accuracy scores of 64% and 67% with the single observing response increased to 79% and 83% with 10 observing responses and increased sample stimulus display durations.

The compound matching task functioned as a DOR in the present experiments. This procedure may also have some diagnostic utility to assess overselective stimulus control in situations in which discrimination of multiple stimuli or multiple stimulus features is important. For example, stimulus control by both a color name and an object name may be needed to accurately identify items. A special-education teacher who suspects overselective control in such discriminations could arrange for the student to identify objects by choosing among a blue cup, red cup, blue ball, and red ball displayed on a shelf. As in the present study, reliably accurate performance under these conditions would verify stimulus control by both the color names and the object names, and a significant number of errors may indicate overselective stimulus control.

Acknowledgments

Research and manuscript preparation was supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development under award numbers R01HD062582 and P30HD004147. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Some of the data were presented at the annual convention of the Association of Professional Behavior Analysts, Seattle, WA (2015) and at the Gatlinburg Conference on Research and Theory in Intellectual and Developmental Disabilities (2015). A portion of the data included in this manuscript was submitted to the Department of Psychology in the School of Arts and Sciences at Western New England University by the first author under the supervision of the second author in partial fulfillment of the requirements for the doctoral degree in behavior analysis. We thank Dr. Gregory Hanley and Dr. Amanda Karsten for their feedback on earlier versions of this manuscript, Eileen Grant, Marlana Mueller, Katherine Nolan, Megan Cicoello, and Josephine Southwick for assistance with data collection and analysis; Lucy Lorin, Jen Brooks, and the Shriver Center's Clinical and Translational Research Support Core for participant recruitment and characterization; and Ben Wallace and Dr. Christophe Gerard for computer software.

References

- Bondy A, Frost L. The Picture Exchange Communication System. *Behavior Modification*. 2001; 25:725–744. doi:10.1177/0145445501255004. [PubMed: 11573337]
- Constantine B, Sidman M. The role of naming in delayed matching to sample. *American Journal of Mental Deficiency*. 1975; 79:680–689. [PubMed: 1146860]
- Cox JK, D'Amato MR. Matching to compound samples by monkeys (*Cebus apella*): Shared attention or generalization decrement? *Journal of Experimental Psychology: Animal Behavior Processes*. 1982; 8:209–225. doi:10.1037/0097-7403.8.3.209.

- Debert P, Matos MA, McIlvane WJ. Conditional relations with compound abstract stimuli using a go/no-go procedure. *Journal of the Experimental Analysis of Behavior*. 2007; 87:89–96. doi: 10.1901/jeab.2007.46-05. [PubMed: 17345953]
- Dickson CA, Deutsch CK, Wang SS, Dube WV. Matching-to-sample assessment of stimulus overselectivity in students with intellectual disabilities. *American Journal on Mental Retardation*. 2006; 111:447–453. doi:10.1352/0895-8017(2006)111[447:MAOSOI]2.0.CO;2. [PubMed: 17029502]
- Doughty AH, Hopkins MN. Reducing stimulus overselectivity through an increased observing-response requirement. *Journal of Applied Behavior Analysis*. 2011; 44:653–657. doi:10.1901/jaba.2011.44-653. [PubMed: 21941399]
- Dube, WV. Stimulus overselectivity in discrimination learning.. In: Reed, P., editor. *Behavioral theories and interventions for autism*. Nova Science Publishers; New York, NY: 2009. p. 23–46.
- Dube WV, Dickson CA, Balsamo LM, O'Donnell KL, Tomanari GY, Farren KM, Wheeler EE, McIlvane WJ. Observing behavior and atypically restricted stimulus control. *Journal of the Experimental Analysis of Behavior*. 2010; 94:297–313. doi:10.1901/jeab.2010.94-297. [PubMed: 21541173]
- Dube WV, Farber RS, Mueller MR, Grant E, Lorin L, Deutsch CK. Stimulus overselectivity in autism, Down syndrome, and typical development. *American Journal on Intellectual and Developmental Disabilities*. 2016; 121:219–235. doi: 10.1352/1944-7558-121.3.219. [PubMed: 27119213]
- Dube WV, McIlvane WJ. Reduction of stimulus overselectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis*. 1999; 32:25–33. doi:10.1901/jaba.1999.32-25. [PubMed: 10201101]
- Dube WV, Wilkinson KM. The potential influence of stimulus overselectivity in AAC: Information from eye-tracking and behavioral studies of attention with individuals with intellectual disabilities. *Augmentative and Alternative Communication*. 2014; 30:172–185. doi: 10.3109/07434618.2014.904924. [PubMed: 24773053]
- Dunn, LM.; Dunn, DM.; Pearson Assessments. PPVT-4: Peabody picture vocabulary test. Pearson Assessments; Minneapolis, MN: 2007.
- Elliott, CD. *Differential Ability Scales*. 2nd ed.. Harcourt Assessment; San Antonio, TX: 2007.
- Farber RS, Dube WV, Dickson CA. A sorting-to-matching method to teach compound matching to sample. *Journal of Applied Behavior Analysis*. 2016; 49:294–307. doi:10.1002/jaba.290. [PubMed: 26843215]
- Gutowski SJ, Stromer R. Delayed matching to two-picture samples by individuals with and without disabilities: an analysis of the role of naming. *Journal of Applied Behavior Analysis*. 2003; 36:487–505. doi:10.1901/jaba.2003.36-487. [PubMed: 14768668]
- Johnson C, Meleshkevich O, Dube WV. Merging separately established stimulus classes with outcome-specific reinforcement. *Journal of the Experimental Analysis of Behavior*. 2014; 101:38–50. doi:10.1002/jeab.61. [PubMed: 24248538]
- Lovaas OI, Koegel RL, Schreibman L. Stimulus overselectivity in autism: A review of research. *Psychological Bulletin*. 1979; 86:1236–1254. doi:10.1037/0033-2909.86.6.1236. [PubMed: 515280]
- Lovaas OI, Schreibman L, Koegel RL, Rehm R. Selective responding by autistic children to multiple sensory input. *Journal of Abnormal Psychology*. 1971; 77:211–222. doi:10.1037/h0031015. [PubMed: 5556929]
- Markham MR, Dougher MJ. Compound stimuli in emergent relations: Extending the scope of stimulus equivalence. *Journal of the Experimental Analysis of Behavior*. 1993; 60:529–542. doi:10.1901/jeab.1993.60-529. [PubMed: 16812718]
- Mayer-Johnson. *Picture Communication Symbols: Boardmaker (Version 6)* [Computer software]. Mayer-Johnson; Solana Beach, CA: 2008.
- McIlvane, WJ. Simple and complex discrimination learning.. In: Madden, GJ., editor. *APA handbooks in psychology. APA handbook of behavior analysis, Vol. 2: Translating principles into practice*. American Psychological Association; Washington, DC: 2013. p. 129–163.

- Ploog BO. Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2010; 40:1332–1349. doi:10.1007/s10803-010-0990-2. [PubMed: 20238154]
- Reed P, Altweck L, Broomfield L, Simpson A, McHugh L. Effect of observing-response procedures on overselectivity in individuals with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*. 2012; 27:237–246. doi:10.1177/1088357612457986.
- Stromer R, McIlvane WJ, Dube WV, Mackay HA. Assessing control by elements of complex stimuli in delayed matching to sample. *Journal of the Experimental Analysis of Behavior*. 1993; 59:83–102. doi:10.1901/jeab.1993.59-83. [PubMed: 8433067]
- Stromer R, McIlvane WJ, Serna RW. Complex stimulus control and equivalence. *Psychological Record*. 1993; 43:585–598.
- Wacker DP, Harding JW, Berg WK, Lee JF, Schieltz KM, Padilla YC, Nevin JA, Shahan TA. An evaluation of persistence of treatment effects during long-term treatment of destructive behavior. *Journal of the Experimental Analysis of Behavior*. 2011; 96:261–282. doi:10.1901/jeb.2011.96-261. [PubMed: 21909168]
- Walpole CW, Roscoe EM, Dube WV. Use of a differential observing response to expand restricted stimulus control. *Journal of Applied Behavior Analysis*. 2007; 40:707–712. doi:10.1901/jaba.2007.707-712. [PubMed: 18189104]
- Yoo JH, Saunders KJ. The discrimination of printed words by prereading children. *European Journal of Behavior Analysis*. 2014; 15:123–135. [PubMed: 25620886]

Figure 1.

Matching-to-sample procedures for pretest and experimental conditions. D-SS = delayed matching with single samples and single comparisons. D-CS = delayed matching with compound (two stimuli) samples and single comparisons. S-CC = simultaneous matching with compound samples and comparisons. DOR = differential observing response procedure with S-CC response embedded within the sample observation portion of a D-CS trial. Note: The images in the figure are not drawn to scale; each stimulus was approximately 1.5×1.5 cm and the monitor display screen was 22.9×30.5 cm.

Figure 2.

Individual session accuracy scores (%) from Experiment 1. A: Data for three representative participants with high D-CS accuracy scores during DOR intervention and accuracy decline with the return to baseline. B: Data for one representative participant with little improvement in D-CS accuracy during DOR intervention. C: Data for one participant with high D-CS accuracy scores during DOR intervention and maintained high accuracy with the return to baseline. D: Data for one representative participant with high D-CS accuracy scores during two DOR interventions and accuracy declines with both returns to baseline.

Figure 3.

Comparison of accuracy scores (%) on DOR titration levels and posttest levels in Experiment 2. Dotted lines show the accuracy criterion of 89%. Data are for seven participants for whom titration estimates accurately predicted the level of DOR support necessary for high accuracy on the posttest.

Figure 4.

Comparison of accuracy scores on DOR titration levels and posttest levels in Experiment 2. Dotted lines show the accuracy criterion of 89%. The top plot shows data for one participant for whom titration results underestimated the level of DOR support necessary for high accuracy on the posttest. The bottom three plots show data for three participants for whom titration results overestimated the level of DOR support necessary for high accuracy on the posttest.

Table 1

Participant Characteristics

Participant	Diagnosis	CA	PPVT MAE	DAS GCA
A30	ID	15.7	4.75	37
A31	ASD	14.3	5.58	40
A34	ID	17.8	8.67	34
A93	ASD	16.4	4.58	45
B06	ASD	11.6	4.75	60
B09	ASD	19.8	4.08	52
B15	ASD	13.8	7.92	51
B18	ASD	18.8	3.17	34
B65	ASD	14.0	2.25	30
B66	ASD	14.9	5.25	30
B68	ASD	15.7	< 2	32
B69	ASD	14.8	3.58	30
B75	ASD	21.0	7.83	51
B78	ASD	13.1	6.75	42
B82	ASD	20.8	2.08	30
B84	ASD	15.1	5.66	54
B89	ASD	12.3	6.50	60
B91	ASD	21.1	2.17	30
B95	ASD	15.3	< 2	30
C09	ID	15.8	6.75	39
C13	ID	14.2	4.08	34
C15	ASD	8.5	4.50	70

Note. ASD = autism spectrum disorder, ID = intellectual disability without ASD. CA = chronological age in years. PPVT MAE = Peabody Picture Vocabulary Test 4 mental-age-equivalent score in years; < 2 indicates untestable. DAS CGA = Differential Abilities Scales II, General Conceptual Ability score (standardized score, $M = 100$, $SD = 15$).

Table 2

Results of Experiment 1, First DOR Intervention

Participant	Stimuli	Baseline 1		DOR Intervention		Baseline 2	
		Sessions	D-CS (%)	S-CC (%)	Remedial Sessions	D-CS (%)	D-CS (%)
A31	Col	8	61	85	8		
B15	Col	11	77	87	11		
B78	Arb	5	71	87	6		
A30	Arb	8	74	98	0	84 (83)	78
A34	Arb	11	79	95	1	92 (94) [*]	86
A93	Arb	14	72	93	4	93 (92) [*]	76
B06	Arb	5	75	100	0	94 (97) [*]	78
B09	Col	5	61	99	0	97 (97) [*]	69
B18	Arb	11	69	99	0	92 (93) [*]	69
B65	Col	8	71	97	0	79 (74)	69
B66	Col	5	73	98	0	90 (89)	77
B68	Arb	11	82	95	1	93 (91)	82
B69	Col	11	69	100	0	96 (99) [*]	71
B75	Arb	5	74	97	0	87 (88)	71
B82	Col	8	72	95	0	89 (87)	80
B84	Arb	8	78	95	0	95 (94) [*]	85 [*]
B89	Arb	8	68	99	0	96 (95) [*]	79
B91	Arb	11	72	92	3	79 (77)	78
B95	Arb	8	73	95	1	87 (87)	79
C09	Col	5	70	97	3	90 (89)	70
C13	Arb	8	68	96	1	91 (90) [*]	66
C15	Arb	5	66	94	4	81 (80)	64

Note. Col = color symbols, Arb = arbitrary forms. Sessions = number of sessions in Baseline 1. D-CS = delayed matching with compound samples and single comparisons. S-CC = simultaneous matching with compound samples and comparisons. D-CS and S-CC accuracy scores are mean percent correct. Baseline 1 accuracy scores calculated from the last five baseline sessions. DOR Intervention D-CS accuracy scores within parentheses are means for the last five intervention sessions.

^{*} Asterisks with D-CS scores indicate mastery criterion (see text for details).

Table 3

Results of Experiment 1, Second DOR Intervention

Participant	Stimuli	Baseline 2	DOR Intervention 2			Baseline 3
		D-CS (%)	S-CC (%)	Remedial Sessions	D-CS (%)	D-CS (%)
A30	Arb	78	98	0	90 (89)*	84
B66	Col	77	99	0	91 (94)*	74
B68	Arb	82	96	1	94 (91)*	80
B75	Arb	71	98	1	91 (93)*	71
B82	Col	80	93	3	91 (92)*	79
B91	Arb	78	97	0	80 (78)	78
B95	Arb	79	98	0	88 (88)	72

Note. Col = color symbols, Arb = arbitrary forms. D-CS = delayed matching with compound samples and single comparisons. S-CC = simultaneous matching with compound samples and comparisons. D-CS and S-CC accuracy scores are mean percent correct. DOR Intervention D-CS accuracy scores within parentheses are means for the last five intervention sessions.

* Asterisks with D-CS scores indicate mastery criterion (see text for details).

Table 4

Experiment 2 Participants and Results

Participant	Total DOR Titration Blocks	Total Remedial Blocks	Estimated DOR Threshold (Levels)
A34	87	4	3-2
A93	80	4	3-2
B06	104	3	5-4
B09	100	0	5-4
B18	82	1	5-4
B66	75	0	5-4
B68	96	1	
B69	138	0	2-1
B75	115	0	
B82	105	4	3-2
B89	75	4	4-3

Table 5

Numbers of Trial Types per DOR Titration Level

Level	DOR	D-CS Baseline (No DOR)	S-CC with Consequences	Total Trials
5	12	0	2	14
4	9	3	2	14
3	6	6	2	14
2	3	9	2	14
1	0	12	2	14
0	0	12	0	12