JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2010, **94,** 297–313 NUMBER 3 (NOVEMBER)

OBSERVING BEHAVIOR AND ATYPICALLY RESTRICTED STIMULUS CONTROL

William V. Dube[1], Chata A. Dickson[1], Lyn M. Balsamo[1], Kristin Lombard O'Donnell[1],
Gerson Y. Tomanari[2], Kevin M. Farren[1], Emily E. Wheeler[1], and William J. McIlvane[1]

[1]UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL - SHRIVER CENTER
[2]UNIVERSIDADE DE SÃO PAULO, BRASIL

Restricted stimulus control refers to discrimination learning with atypical limitations in the range of controlling stimuli or stimulus features. In the study reported here, 4 normally capable individuals and 10 individuals with intellectual disabilities (ID) performed two-sample delayed matching to sample. Sample-stimulus observing was recorded with an eye-tracking apparatus. High accuracy scores indicated stimulus control by both sample stimuli for the 4 nondisabled participants and 4 participants with ID, and eye tracking data showed reliable observing of all stimuli. Intermediate accuracy scores indicated restricted stimulus control for the remaining 6 participants. Their eye-tracking data showed that errors were related to failures to observe sample stimuli and relatively brief observing durations. Five of these participants were then given interventions designed to improve observing behavior. For 4 participants, the interventions resulted initially in elimination of observing failures, increased observing durations, and increased accuracy. For 2 of these participants, contingencies sufficient to maintain adequate observing were not always sufficient to maintain high accuracy; subsequent procedure modifications restored it, however. For the 5th participant, initial improvements in observing were not accompanied by improved accuracy, an apparent instance of observing without attending; accuracy improved only after an additional intervention that imposed contingencies on observing behavior. Thus, interventions that control observing behavior seem necessary but may not always be sufficient for the remediation of restricted stimulus control.

Key words: observing behavior, restricted stimulus control, stimulus overselectivity, matching to sample, touchscreen, humans

Discrimination training may result in stimulus control by only a subset of the potential controlling stimuli (e.g., Reynolds, 1961). A number of factors may be related to restrictions in the range, breadth, or number of controlling stimuli. In normally capable humans, these factors include the number of stimuli or relevant stimulus features, increase in task difficulty, requirements for rapid responding, or some concurrent and distracting activity (e.g., Critchfield & Perone, 1993; Dube et al., 2006; McHugh & Reed, 2007). However, if the breadth of stimulus control is atypically limited, the outcome has been termed *restricted stimulus control* (Dube & McIlvane, 1997; Litrownik, McInnis, Wetzel-Pritchard, & Filipelli, 1978) or stimulus overselectivity (reviewed in Lovaas, Koegel, & Schreibman, 1979; and Dube, 2009). Atypically restricted stimulus control is often observed in individuals who have intellectual or neurodevelopmental disabilities such as autism spectrum disorders, and it is a widely acknowledged problem in the education of such individuals (e.g., Barthold & Egel, 2001; Bickel, Richmond, Bell, & Brown, 1986; Schreibman, 1997). For example, special-education students may identify printed words or other similar arrays of characters on the basis of the initial letter only (e.g., Dickson, Wang, Lombard, & Dube, 2006).

This research was supported by Grant Numbers P30 HD004147, P01 HD025995 and R01 HD037055 from the National Institute of Child Health and Human Development. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of NICHD.

A detailed analysis of the data from Participant DTM's first baseline session was previously published in Dube et al. (1999).

Chata Dickson is currently at The New England Center for Children.

For their generosity with technical support and assistance, we thank Rikki Razdan and Alan Kielar of ISCAN Inc., and Karen DeGregory and David Pasterchik of Abilities Software. Special thanks to Tom Fowler for assistance in data collection and contributions to the development of data analysis protocols. For assistance in data collection, we also thank Lauren Abraham, Jennifer Brooks, Claudia Heye, Susan Kelledy, Jason Krienke, Sarah Luthern, Rena Sphritzer, Brooks Thompson, and Sharon Wang. We thank the staff and students of the New England Center for Children and the Protestant Guild Learning Center for their cooperation.

Address correspondence to William V. Dube, UMMS Shriver Center, 333 South Street, Shrewsbury, MA 01545 (e-mail: william.dube@umassmed.edu).

doi: 10.1901/jeab.2010.94-297

When considering restricted stimulus control, a question arises about the role of observing behavior. For visual stimuli, effective observing responses include head movements and eye orientations that cause light reflected by the stimulus to fall upon the retina. On the one hand, it is possible that the individual who displays restricted stimulus control has observed all of the relevant stimuli. If so, then one might conclude that the deficiency is related to attending (see Dinsmoor, 1985, for comments on the distinction between observing and attending). On the other hand, it is possible that the individual has not observed all of the relevant stimuli. In this case, restricted stimulus control might be attributed to inadequate contact with the stimuli and some intervention that corrected this inadequacy in observing behavior would be required before any analysis of attending could proceed.

Dube et al. (1999) evaluated restricted stimulus control in one normally capable adult and one 12-year-old boy with moderate intellectual disability (ID). The discrimination procedure was delayed matching-to-sample (DMTS) with two sample stimuli per trial, displayed side-by-side. On each trial, only one of the samples appeared in the comparison stimulus array, and thus high accuracy required observing and attending to both samples on every trial (details in the Methods section below). Observing behavior was measured for one session with an eye-tracking apparatus. The adult's accuracy score was high (96%) and a trial-by-trial analysis of observing behavior while the sample stimuli were displayed showed a regular back-and-forth pattern that included at least one observation of each sample on every trial. The boy with ID exhibited intermediate accuracy (61%) consistent with restricted stimulus control, his observing patterns were highly variable, and he failed to observe one of the samples on many trials. The first goal of the study reported here was to expand the data base of these findings by examining the relations between observing behavior and two-sample DMTS accuracy in 14 participants, 4 normally capable individuals and 10 with ID.

The second goal of the study concerned those participants who had intermediate accuracy scores and observing failures during the initial evaluation. Would changes in the experimental procedures that eliminated observing failures and increased observing durations produce high accuracy scores and thus eliminate restricted stimulus control? Adequate contact with sensory receptors seems a necessary prerequisite for effective performance and previous research in intellectual and neurodevelopmental disabilities has shown that some performance deficits may be corrected by identifying and supplying missing prerequisites (e.g., Dube, Iennaco, Rocco, Kledaras, & McIlvane, 1992; Sidman & Stoddard, 1966).

To address this second goal, we were concerned primarily with answering the question of whether or not the individuals with intermediate accuracy were in fact capable of exhibiting higher accuracy. Individuals with neurodevelopmental disabilities have often been characterized as showing great heterogeneity in response to intervention (e.g., Howlin, Magiati, & Charman, 2009; Thorsteinsson et al., 2007), thus leaving open the issue of whether restricted stimulus control can be corrected via training. It seems possible that restricted stimulus control might reflect neurological inadequacy (e.g., Allen & Courchesne, 2003; Belmonte & Yurgelun-Todd, 2003). In addition, changing the topography of peripheral behavior such as eye movements may have little effect on performance if the deficit is due to other behavior that may occur independently of the eye movements. Several procedures were employed as interventions to change observing behavior, alone or in combination: differential reinforcement for longer observing durations, extra-stimulus prompts, within-stimulus prompts, a contingency that extended the sample observation period until observing occurred, and, in one case, increased reinforcement for high accuracy scores. Based on experience with other manipulations intended to broaden restricted stimulus control (e.g., Dube & McIlvane, 1999), we anticipated that procedures that proved effective with some participants might not be effective with others. Thus, a number of procedure alternatives were made available. The general strategy was to try one or more interventions until one was effective in that it eliminated observing failures, increased observing durations to exceed those of the high-accuracy participants, and increased accuracy scores to at least 85%. After a brief exposure to an effective intervention, it was then withdrawn and baseline conditions reinstated to demonstrate control by the intervention.

We emphasize that the intervention procedures were intended to correct inadequate observing behavior by any effective arrangement of stimulus manipulation or reinforcement contingencies. The goal was to produce high accuracy and observing behavior with frequency and duration characteristics that met or exceeded those of high-accuracy participants. The experiment was not designed to determine the most efficient version of any individual remedial procedure, nor was it designed to compare the relative effectiveness of different procedures. Thus, procedural controls necessary for evaluations of relative efficiency and effectiveness were not included in this study.

METHOD

Participants, Setting, and Apparatus

Participants were 4 adults with no known clinical condition, recruited through personal contacts, and 10 individuals with ID attending residential schools. Participants with ID were included in this study if they were able to perform simple matching-to-sample tasks (details below) and follow simple verbal instructions. Table 1 shows participant characteristics.

Experimental sessions were conducted in two laboratories equipped with similar apparatus. One testing room was 4 × 2.5 m and the other was 3.4 × 1.5 m. On one side of each room was a chair for the participant, a small table with a computer, a color monitor with a touch-sensitive screen, and an automatic token dispenser (Med ENV-703). The computer was used to display stimuli to the participant. On the other side of each room was a chair for the experimenter and a table with the eye-tracking apparatus. During sessions, a curtain was drawn across the room to separate the participant from the experimenter and eye-tracking apparatus.

The eye tracking apparatus consisted of an ISCAN RK-426PC Pupil/Corneal Reflection Tracking System, RK520PC Autocalibration System, and Miniature Head-Mounted Eye Imaging and Line-of-Sight Scene Imaging Systems (ISCAN Inc., Burlington, MA). These systems were integrated with a desktop computer running ISCAN Point of Regard Data Acquisition and Fixation Analysis software. The participant wore an adjustable plastic headband with two miniature video cameras attached by supporting arms. One camera produced an image of the participant's eye, and the other camera produced an image of the central portion of the participant's field of view from the participant's perspective. Because the imaging components were head-mounted, the distance and angle between the cameras and the participant's eye did not change as a result of head movement, and thus it was not necessary to immobilize the head during recording.

The apparatus performed on-line image analysis and produced a real-time video image showing the central portion of the participant's field of view with a superimposed cursor indicating the location of the participant's visual gaze (see Dube et al., 1999 for additional apparatus details). Vertical interval time code (VITC) was added to the video signal and this composite video image was displayed on a monitor in the experimenter's area of the laboratory and recorded on videotape cassettes.

Two-Sample Delayed Matching-to-Sample (DMTS) Procedure

Stimuli were black abstract forms, approximately 1.25 cm square, displayed on a white background. The stimuli were similar to Greek letters or dingbat font characters (for examples see Dube & McIlvane, 1997). Throughout the study, different stimuli appeared on every trial within each session. The stimuli were drawn at random, without replacement, from a pool of 180 different forms.

Trials began when two sample stimuli appeared side-by-side in the center of the screen, 4 cm center to center (4.2° visual angle at a viewing distance of 55 cm). When the participant touched the sample display area, the sample stimuli disappeared from the screen. After a delay of 1 s (for normally capable adults) or 0 s (for participants with ID), three comparison stimuli appeared in three corners of the screen; the location of the blank corner varied from trial to trial. One comparison was identical to one of the samples, and the other two comparisons did not match either of the samples. Touching the identical comparison was the correct response and touching either of the nonidentical

Table 1
Participant Characteristics and Duration of Initial Training to Use Eye-Tracking Apparatus.

Participant(s)	Diagnosis	Age	Peabody Picture Vocabulary Test-R Mental Age Equivalent Score	Number of Eye Tracking Training Sessions
JNB, NLI, RBO, GGN		23–31		[a]
JAM	Mild Intellectual Disability, Post Traumatic Stress Disorder	17	12.1	1
CHR	Mild-Moderate Intellectual Disability, Attention Deficit Hyperactivity Disorder	17	7.7	1
FAB	Mild-Moderate Intellectual Disability, Behavior Disorder	19	9.1	1
AAT	Developmental Delay	17	9.1	1
TRZ	Mild Intellectual Disability, Emotional Disorder	21	10.4	2
DDA	Mild-Moderate Intellectual Disability, Behavior Disorder	12	5.3	14
DTM	Mild-Moderate Intellectual Disability	12	4.2	16
WLN	Cerebral Palsy, Intellectual Disability	14	6.1	5
STN	Mild-Moderate Intellectual Disability	16	8.6	23
MAR	Pervasive Developmental Delay	17	6.3	6

[a] Normally capable participants received verbal instructions for eye tracking calibration at the beginning of the experimental session.

comparisons was an error. The stimulus in the left or right sample position was correct equally often and correct sample position varied from trial to trial in unsystematic order. During the sample display period, therefore, participants could not predict which sample would be the correct comparison. There was a 3-s intertrial interval (ITI) with a blank display screen.

Consequences

For the normally capable participants, all correct responses were followed by a 0.5-s beep, an increment of a 1 × 2 cm counter displayed in the top center of the monitor screen, and the ITI. Points were exchanged for money after sessions; participants received approximately $15 per session. For the participants with ID, all correct responses were followed by a 1.5-s auditory-visual computer display of animated stars, melodic tones, an automatically dispensed token, and the ITI. After sessions participants exchanged the tokens for their choices of snack foods, toys, brief activities, or money. For all participants, trials that ended with errors were followed only by the ITI.

General Procedure

Preliminary training. Participants with ID were given pretests to verify that they could perform (a) a one-sample version of the DMTS procedure on which only one sample stimulus appeared on each trial, and (b) a simultaneous-matching version of the two-sample procedure on which the sample stimuli remained displayed throughout the trial. All participants were familiar with matching-to-sample tasks and all had accuracy scores of at least 90% in 36-trial sessions with each procedure. These participants were also trained to accumulate plastic poker-chip tokens for correct responses on the matching task and exchange the tokens for other items after the session.

Participants with ID were trained to wear a plastic replica of the eye tracking headgear and then to participate in a brief calibration routine using the actual headgear (details below). Details of the training program appear in Dube et al. (1999). The number of eye tracking training sessions for each participant with ID is shown in the rightmost column of Table 1.

Experimental sessions. Participants were seated in front of the stimulus-presentation computer monitor at a viewing distance of approximately 55 cm and the imaging headgear was adjusted by an experimenter. Each session began with a brief calibration routine in which participants were instructed to hold the head still and fixate on targets that appeared in various locations on the stimulus display monitor. For participants with ID, an experimenter stood behind the participant and positioned his/her head during calibration. Calibration targets for participants with ID included 2 × 2 cm pictures or short video clips. To control observing behavior, participants were asked to name the pictures aloud or comment on the video clips. After calibration, all participants were told that they need no longer hold the head still.

For normally capable adults a window appeared on the monitor screen with the following instructions: ‘‘In this experiment, shapes will appear on the computer screen. You can earn points by touching the shapes. The window at the top of the screen will show your score. Try to get the highest score you can. Touch ‘Continue’ to begin.’’ Participants were asked to read the instructions aloud, and then they touched an on-screen button labeled ‘‘Continue’’ to begin the session.

Baseline Evaluation of Two-sample DMTS

All participants were given a *Baseline* evaluation with the two-sample DMTS procedure described above. For all participants except MAR, sessions consisted of 3 two-sample simultaneous matching trials followed by 36 two-sample DMTS trials. MAR’s sessions included 24 two-sample DMTS trials (to reduce her session duration). Participants with accuracy scores above 90% received one session. Participants with lower accuracy scores received three sessions with three exceptions: Participant TRZ withdrew from the experiment after two sessions because he expressed unwillingness to wear the eye tracking headgear, Participant STN was given four sessions because the eye tracking signal was lost for approximately half of the trials in two of her sessions, and Participant MAR received five sessions because of the smaller number of trials per session.

Intervention Procedures

Five participants who completed the Baseline evaluation with intermediate accuracy scores were given one or more interventions designed to increase observing behavior, followed by at least one return to Baseline conditions. The initial intervention for each participant was one of the first three described below (differential reinforcement, extra-, or within-stimulus prompts), employed initially because these types of procedures are commonly used with individuals who have developmental disabilities (e.g., Lancioni & Smeets, 1986). If the initial intervention was not followed by increases in observing frequency, observing duration, and accuracy, then additional behavioral contingencies were imposed (described below), either alone or in combination with a prompting procedure.

Differential reinforcement for observing. The trial procedure was the same as the Baseline condition. During trials the experimenter watched the video image displayed on the eye-tracking monitor in the experimenter’s area of the laboratory and controlled the token dispenser with a hand switch. On trials with a correct response that followed sample observing durations of at least 0.5 s (estimated by the experimenter) for both left and right sample stimuli, the experimenter delivered four tokens. On trials with a correct response but with one or both sample observing durations less than 0.5 s, the experimenter delivered only one token.

Extrastimulus prompts. Extrastimulus prompts consisted of the experimenter’s vocalizations or vocalizations plus gestures. *Vocal prompts* included phrases such as ''Look at this one, that one, this one, that one,'' or ''… now the other one, now the first one, the other one,'' and so forth. The experimenter watched the eye-tracking monitor and delivered prompts until the participant completed a total of four observations with estimated durations of at least 0.5 s, two each of the left and right samples. When vocal prompts were accompanied by *gestural prompts,* the experimenter pointed to a sample stimulus as he or she delivered the vocal prompt. The experimenter's fingertip was approximately 1 cm above the stimulus during a gestural prompt.

Within-stimulus prompts. Observing was prompted by abrupt changes in the color, or color and size, of the sample stimuli. The rationale was that a series of sudden changes in salience differences between the sample stimuli may exert control of observing (Theeuwes, 1994; Yantis, 1996). During the ITI before each trial, the experimenter watched the eye-tracking monitor and, when the position of the point-of-gaze cursor indicated that the participant was looking at the stimulus display screen, pressed a key to terminate the ITI and present the sample stimuli. For *within-stimulus color prompts,* the sample stimulus on the left side was initially presented in black (as usual) and the sample on the right side in a light gray color. For *within-stimulus color and size prompts,* the black stimulus was presented at 175% of its normal size and the gray stimulus at the normal size. After 1 s, the color or color/size changes reversed, so that the left-side sample became light gray and normal size, and the right-side sample became black and also larger if size prompts were included. The color or color/size changes continued to reverse at 1 s intervals for a total of 6 s. During the prompting sequence, responses to the touchscreen had no effect. When the prompting sequence was complete, the touchscreen became active and the trial proceeded according to the Baseline procedure.

Participant STN received some sessions in which the color and size prompts were faded. In *Fading Step 1,* the larger stimulus increased to 130% or 150% of normal size, the smaller stimulus was presented in a medium gray color, and the sequence duration was 4 s. In *Fading Step 2,* the larger stimulus was 125% normal size, the smaller stimulus was dark gray, and the sequence duration was 4 s.

Observing contingency. A contingency on observing behavior was imposed by the experimenter. When the sample stimuli were presented, the touchscreen became inactive. The experimenter observed the eye-tracking monitor until there were at least two fixations on each sample stimulus, each at least 0.5 s in duration (estimated), and alternating between the left and right sample positions. The experimenter then activated the touchscreen by pressing a button and the trial proceeded according to the Baseline procedure. Thus, the contingency specified that the trial could not advance until there were at least four sample observations and a total observing duration of 1 s for each sample stimulus.

Within-stimulus prompts plus observing contingency. In some sessions the within-stimulus prompt and observing contingency procedures described above were combined. The experimenter controlled the duration of each of the successive larger/black versus smaller/gray sample displays by pressing a key. The experimenter watched the eye-tracking monitor and did not press the key to advance the prompt sequence until the point-of-gaze cursor indicated a fixation of at least 0.5 s (estimated) on the larger sample stimulus and thus at least six sample observations and total observing duration of 1.5 s for each sample stimulus.

High-accuracy contingency. For Participant STN only, the reinforcement contingencies were modified to add a monetary bonus for accurate performance (rationale described with the results). At the end of each session with the high-accuracy contingency, the computer screen displayed the following message: ''Your score is [number of correct trials]. [participant's name], did you earn 33 points or more? If you did, ask [experimenter's name] for an extra $1.00.''

RESULTS

Baseline Evaluation of Two-sample DMTS

Accuracy scores. The ''Session Accuracy'' column in Table 2 shows overall two-sample DMTS accuracy scores during initial Baseline testing; mean accuracy is shown for those participants who received more than one session. The upper portion of Table 2 shows

Table 2

Initial Baseline Accuracy Scores and Characteristics of Observing Behavior from Eye-Tracking Data Analyses.

Participant	Session Accuracy (%)	Trials Coded	Coded Trials Accuracy (%)[a]			Duration per Trial (s)			Frequency per Trial	
			All	L	R	L	R	N	L	R
JNB	100	26	100	100	100	1.18	.84	.04	2.04	1.77
NLI	94	25	92	93	91	.66	.77	.03	1.92	1.36
RBO	94	26	96	100	92	.76	.74	.03	2.35	1.81
GGN	92	26	92	92	93	.34	.47	.01	1.61	1.00
JAM	100	26	100	100	100	.86	.39	.01	1.38	1.38
CHR	100	26	100	100	100	.65	.44	.02	1.81	1.42
FAB	97	25	100	100	100	1.01	.49	.19	2.12	1.60
AAT	92	77	92	95	89	.89	.64	.13	1.39	1.11
TRZ	60	48	63	58	67	.40	.49	0	.71	.85
DDA	76	63	78	94	64	.51	.17	.17	1.18	.60
DTM	61	72	60	37	86	.42	.90	.56	.97	1.51
WLN	68	78	68	77	58	.89	.31	.02	1.36	.94
STN	72	82	65	36	93	.15	.54	.04	.48	1.09
MAR	68	67	64	82	45	.48	.21	.08	1.01	.84

[a] Coded trials accuracy scores are shown for All trials, and for trials on which the correct comparison was identical to the left (L) or right (R) sample stimulus.

that accuracy scores for all 4 normally capable adults and 4 participants with ID were above 90%. Participant AAT's mean accuracy score was 92%; she received three sessions because the accuracy score in her first session was 89%. The lower portion of Table 2 shows that 6 participants with ID had two-sample DMTS accuracy scores in the intermediate range (60% to 76%) indicative of restricted stimulus control.

Analysis of eye-tracking data. Eye-tracking videotapes were analyzed with OCS Tools software (Research Triangle Collaborative, Raleigh, NC) or Video Frame Coder software (Abilities Software, Sudbury, MA). These software programs facilitated frame-by-frame coding of the video recordings; the frame rate was 30 frames per second. To increase the probability of sampling stable performances, the analyses did not include the first 10 two-sample DMTS trials from each session. Eye-tracking signals were occasionally missing because of events such as movement of the headgear on the head, pronounced squinting, or significantly skewed head and eye positions. Trials were not included in the analyses if the signal was missing for more than 50% of the time in which the sample stimuli were visible within the video frame. The ''Trials Coded'' column in Table 2 shows the total number of trials coded for each participant. The ''Coded Trials Accuracy, All'' column shows the accuracy scores for those trials included in the eye-tracking data analyses. Accuracy scores for the coded trials were similar to the accuracy scores for the entire sessions in all cases. With respect to accuracy, therefore, the eye-tracking samples were representative of each participant's performance.

An *observation* of a sample stimulus was defined as occurring when (a) the sample stimuli were displayed and (b) the center of the point-of-gaze cursor was within a target area surrounding each sample stimulus. Target-area boundaries were derived from the point midway between sample stimuli, such that the stimulus was in the center of each target area and the border between the target areas was halfway between the stimuli. During those portions of the videos when sample stimuli were displayed, video frames were assigned one of four possible codes: Left, Right, Neither, or No Signal. If the point-of-gaze cursor was very close to a target area boundary, the experimenter displayed the individual video frames on a computer monitor and used a millimeter ruler to measure the relevant distances. If this measurement showed that the cursor was exactly centered on a boundary, the frame was assigned the same code as the previous frame.

To evaluate coding accuracy, a second experimenter independently coded one initial Baseline session for each participant. For those

Table 3

Intercoder Agreement for Eye-Tracking Data.

Participant	Condition	Agree (frames)	Disagree (frames)	Percent Agreement
JNB	Baseline	1548	93	94.3%
NLI	Baseline	1031	140	88.0%
RBO	Baseline	1118	86	92.9%
GGN	Baseline	606	43	93.4%
JAM	Baseline	875	22	97.5%
CHR	Baseline	813	45	94.8%
FAB	Baseline	1061	99	91.5%
AAT	Baseline	1117	46	96.0%
TRZ	Baseline	651	19	97.2%
DDA	Baseline	562	37	93.8%
	Within-stimulus prompts	3348	203	94.3%
	Baseline	595	45	93.0%
DTM	Baseline	773	81	90.5%
	Within-stimulus prompts + observing contingency	13216	1242	91.4%
	Baseline	1297	38	97.2%
WLN	Baseline	944	44	95.5%
	Within-stimulus prompts + observing contingency	6056	375	94.2%
	Baseline	1234	90	93.2%
STN	Baseline	433	30	93.5%
	Within-stimulus prompts	6379	199	97.0%
	Baseline	607	40	93.8%
MAR	Baseline	553	65	89.5%
	Extra-stimulus prompts	1565	149	91.3%
	Baseline	645	73	89.8%
Total		47027	3304	93.4%

participants who received interventions, intercoder reliability was also assessed for one intervention session in which accuracy was high, and also for one subsequent Baseline session. Table 3 shows the number of video frames in which both coding records were coded for an observation of the same stimulus (Agreements, either both Left, or both Right), the number of frames in which at least one record was coded for observation of a stimulus (either Left or Right) but the two records did not agree (Disagreements, e.g., Left in one record and Neither in the other record), and the percent agreement calculated by dividing agreements by the sum of agreements plus disagreements. The bottom row of Table 3 shows that the overall percent agreement score was 93.4% (range in individual sessions 88.0% to 97.5%). Almost all disagreements occurred on video frames in which the point-of-gaze cursor appeared to be on a target area boundary, for example, as it was moving from the Left stimulus to the Right stimulus.

General characteristics of observing behavior. The ''Duration'' columns in Table 2 show mean observing durations per trial. Durations for left (L) and right (R) sample stimulus positions were calculated by dividing the total time during which the point-of-gaze cursor was within the target areas for each sample stimulus position by the number of trials. The longest mean observing durations were found in participants with high accuracy (1.18 and 1.01 s; Participants JNB and FAB), and the shortest durations were found in participants with intermediate accuracy (.15 to .21 s; STN, DDA, MAR). In general, however, there was considerable overlap in the distribution of observing durations between high- and intermediate-accuracy participants.

Most participants had asymmetries in observing durations for left versus right sample stimulus positions, and the asymmetries were more pronounced in the participants with ID. All participants with intermediate accuracy scores and two with high accuracy (JAM and FAB) had asymmetries with disparities of 2:1 or greater. The ''Coded trials accuracy, L'' and ''R'' columns show accuracy scores for trials on which the correct comparison was identical to the stimulus in the left or right sample position, respectively. For the participants who had intermediate accuracy, Table 2 shows that left/right asymmetries in observing durations were correlated with left/right accuracy differences. One apparent exception is Participant

TRZ, who was the only participant whose sample-position observing asymmetry changed across sessions. TRZ favored the left sample position in his first session and the right sample position in his second. All other participants with asymmetries consistently observed one sample position for longer mean durations than the other position. The individual session data show that TRZ's asymmetries in observing durations were also correlated with accuracy differences: In his first session, mean observing durations and accuracy scores were .65 s and 67% for the left position and .27 s and 50% for the right. In his second session, durations and accuracy were .15 s and 50% for the left and .71 s and 83% for the right.

The column labeled ''Duration per trial, N'' shows the per-trial duration during which neither stimulus was observed during observing bouts. This event occurred when the point-of-gaze cursor moved from within a sample stimulus target area to a point that was not within either target area (any time between the onset of the sample stimuli and the initial observation of one of the samples was not included). That is, the participant who had been observing a stimulus looked away from both stimuli. These ''Neither'' durations were relatively brief for participants with high accuracy and 4 of 6 participants with intermediate accuracy, and they indicate that these participants rarely looked away from the sample stimuli. In 2 participants with intermediate accuracy, DDA and DTM, the durations for looking away were within the range of durations for observing sample stimuli, that is, looking-away durations were comparable to sample-observing durations.

The ''Frequency per Trial'' columns in Table 2 show the mean number of left and right sample position observations per trial, calculated by dividing observing totals for each position by the number of trials. For example, a left–right–left observing sequence was scored as two observations of the left stimulus and one observation of the right stimulus. In calculating frequency, any successive observations of the same stimulus (interrupted by looking away from both stimuli) were collapsed; for example, left–right–neither–right was scored as one observation of the left and one of the right. All participants with high accuracy had mean observing frequencies of at least one per stimulus position. All participants with intermediate accuracy had a frequency of less than one for at least one stimulus position. Thus, intermediate accuracy was accompanied by failures to observe all of the stimuli.

Detailed analysis of observing frequency. Figure 1 presents a more detailed picture of observing frequencies. The figure shows frequency distributions for the number of observations per trial for each participant. For example, a left–right–left observing sequence was scored as three observations. The top row of Figure 1 shows that the normally capable adults made three, four, or five observations on most trials, and rarely or never made only one observation. (One observation indicates that one of the two sample stimuli was not observed.) The second row of Figure 1 shows that the 4 participants with ID and high accuracy made two, three, or four observations on most trials and rarely had trials with only one observation, with the exception of Participant AAT who did so on 22% of trials.

The bottom two rows of Figure 1 show that all participants with ID and intermediate accuracy scores made only one observation on a substantial proportion of trials (range: 32% [WLN] to 60% [TRZ]). Thus, intermediate accuracy scores were accompanied by incomplete observing of the sample stimulus array.

Table 4 shows the relation between incomplete observing and accuracy scores. The ''Number of Trials with One Observation'' column shows the total number of trials on which the participant observed only one sample stimulus. The next three columns, labeled ''Observed Correct Sample'' show data for those trials on which the single sample stimulus that was observed subsequently appeared in the comparison array as the correct comparison. All participants had very high accuracy scores on these trials. The last three columns, labeled ''Failed to Observe Correct Sample'' show data for trials on which the single sample stimulus that was observed did not appear in the comparison array. All participants' accuracy scores were low (range 13% to 54%) on these trials. These data show that the intermediate accuracy scores for trials with one observation resulted from averaging high-accuracy trials on which the participant had observed the correct sample stimulus with low-accuracy trials on which the participant did not observe the correct sample stimulus.

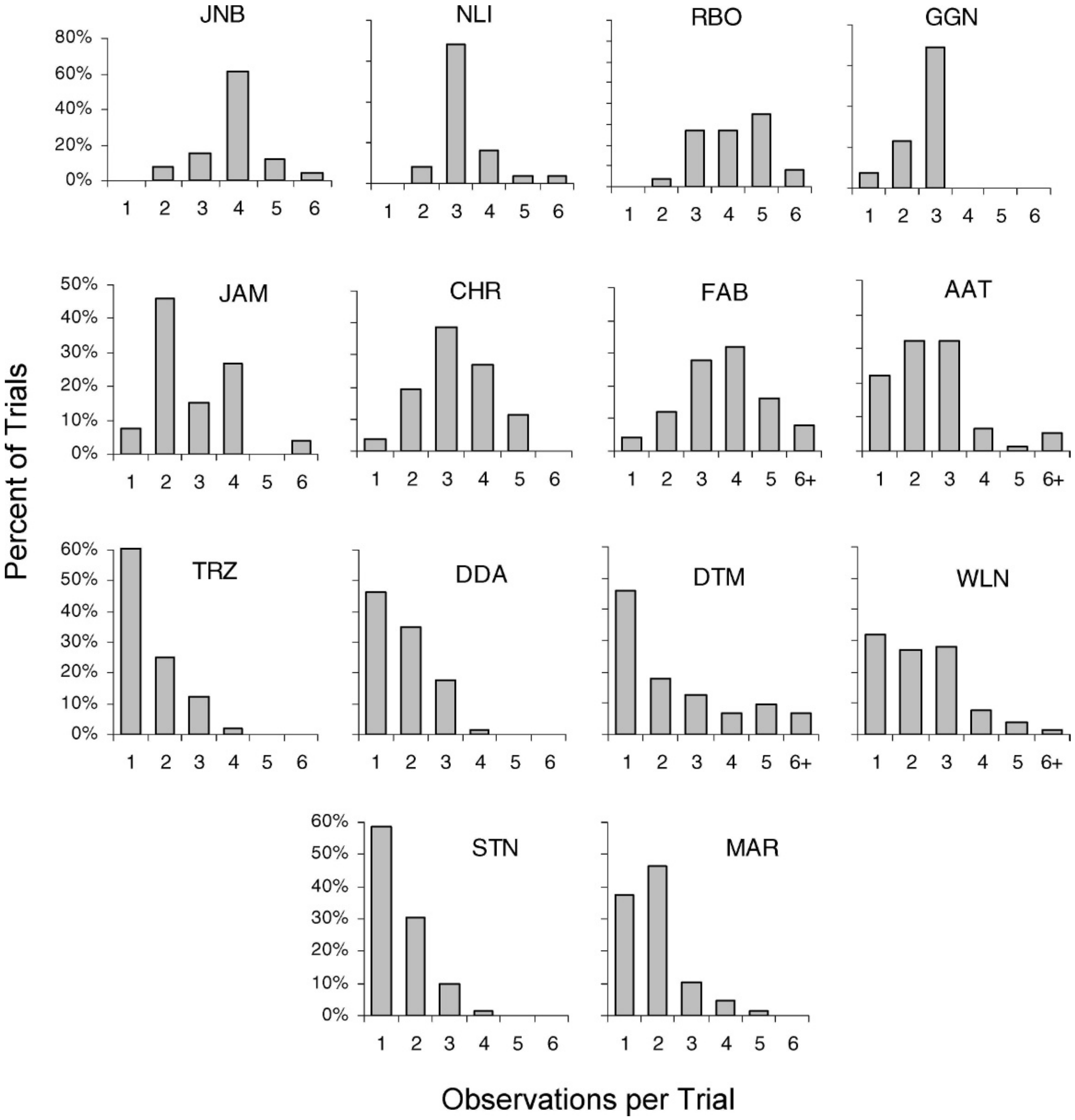

Fig. 1. Distribution of number of observations per trial during initial Baseline sessions. Rightmost columns labeled "6+" show percent of trials with six or more observations.

Detailed analysis of observing duration. Trials with only one observation do not account for all of the errors. All participants with intermediate accuracy scores also made errors on trials in which they had observed both sample stimuli. Therefore, there were some instances in which a recorded observation of a stimulus was not followed by stimulus control.

An additional analysis was conducted to examine the relation between trial outcome and duration observing the sample stimulus that subsequently appeared as the correct comparison stimulus (recall that the participant could not predict which sample would be correct). Figure 2 shows observing durations for the correct sample stimulus on trials that subsequently ended with correct responses (distribution on the left within each panel) or errors (distribution on the right). Solid and dashed horizontal lines within each distribu-

Table 4

Initial Baseline Trials with One Observation: Observing Correct Sample Stimulus and Accuracy Scores.

Participant	Number of Trials with One Observation	Observed Correct Sample			Failed to Observe Correct Sample		
		Correct Trials	Error Trials	Accuracy	Correct Trials	Error Trials	Accuracy
TRZ	29	14	1	93%	5	9	36%
DDA	29	14	0	100%	7	8	47%
DTM	33	15	0	100%	3	15	17%
WLN	25	12	0	100%	7	6	54%
STN	48	19	1	95%	9	19	32%
MAR	25	10	0	100%	2	13	13%

tion show mean and median durations, respectively. Although there was substantial variability, mean and median observing durations for all participants were greater on trials with correct responses than trials with errors. The magnitude of the difference in means varied from a ratio of 1.18 (duration correct / duration error) for TRZ to 2.29 for DTM.

Interpretation of the data shown in Figure 2 is complicated by the difference between the number of observed errors and the number of stimulus-control failures (e.g., Fetterman, 1991). In the three-choice DMTS task, one may assume that one-third of all responses not controlled by a sample stimulus would be scored as correct ''by chance.'' A conservative estimate, therefore, is that the number of observed errors is only two-thirds of the number of stimulus-control failures. Thus, in Figure 2 the distributions for trials with correct responses include some trials on which the controlling stimulus was not the sample (i.e., correct by chance), and these trials are not included in the distributions for trials with errors. If the durations on trials correct by chance were comparable to those on trials with errors, then the net effect for the data as shown in Figure 2 would be to depress the mean and median durations for correct responses and increase the variability of those distributions. It is impossible, however, to identify the specific trials that were correct by chance.

Despite the generally shorter observing durations on trials with errors, duration alone is not a sufficient explanation for the errors. For example, Figure 2 shows that participants also made one (DDA), two (STN, WLN), or three (TRZ, DTM, MAR) errors on trials after observing the correct sample stimulus for a duration that was greater than the mean duration for correct responses. The data support a conclusion only that longer observing durations were somewhat more likely to be accompanied by stimulus control.

To summarize the results of the initial Baseline condition: (a) All normally capable adults and 4 participants with ID (40%) had high accuracy scores on the two-sample DMTS task; 6 participants with ID (60%) had intermediate accuracy scores consistent with restricted stimulus control. (b) Participants with longer observing durations tended to have higher accuracy scores. (c) Asymmetries in duration observing the left versus right sample position were greater in participants with intermediate accuracy and the asymmetries were related to accuracy scores. (d) Failures to observe one sample stimulus occurred almost exclusively in participants with intermediate accuracy and these failures were directly related to errors. (e) Participants with intermediate accuracy also made some errors on trials in which both sample stimuli were observed, and errors tended to follow shorter observing durations on these trials. (f) On approximately 3% of trials, participants with intermediate accuracy observed the correct sample stimulus for durations typical of accurate performance but made errors nevertheless.

Intervention Procedures

Table 5 shows the intervention results. For each participant, the results of the first Baseline condition (in Table 2 and Figure 1) are included in Table 5 to facilitate evaluation of the effects of the interventions. For each condition, Table 5 shows the number of sessions, the mean accuracy score, and three measures of observing behavior: the percent of

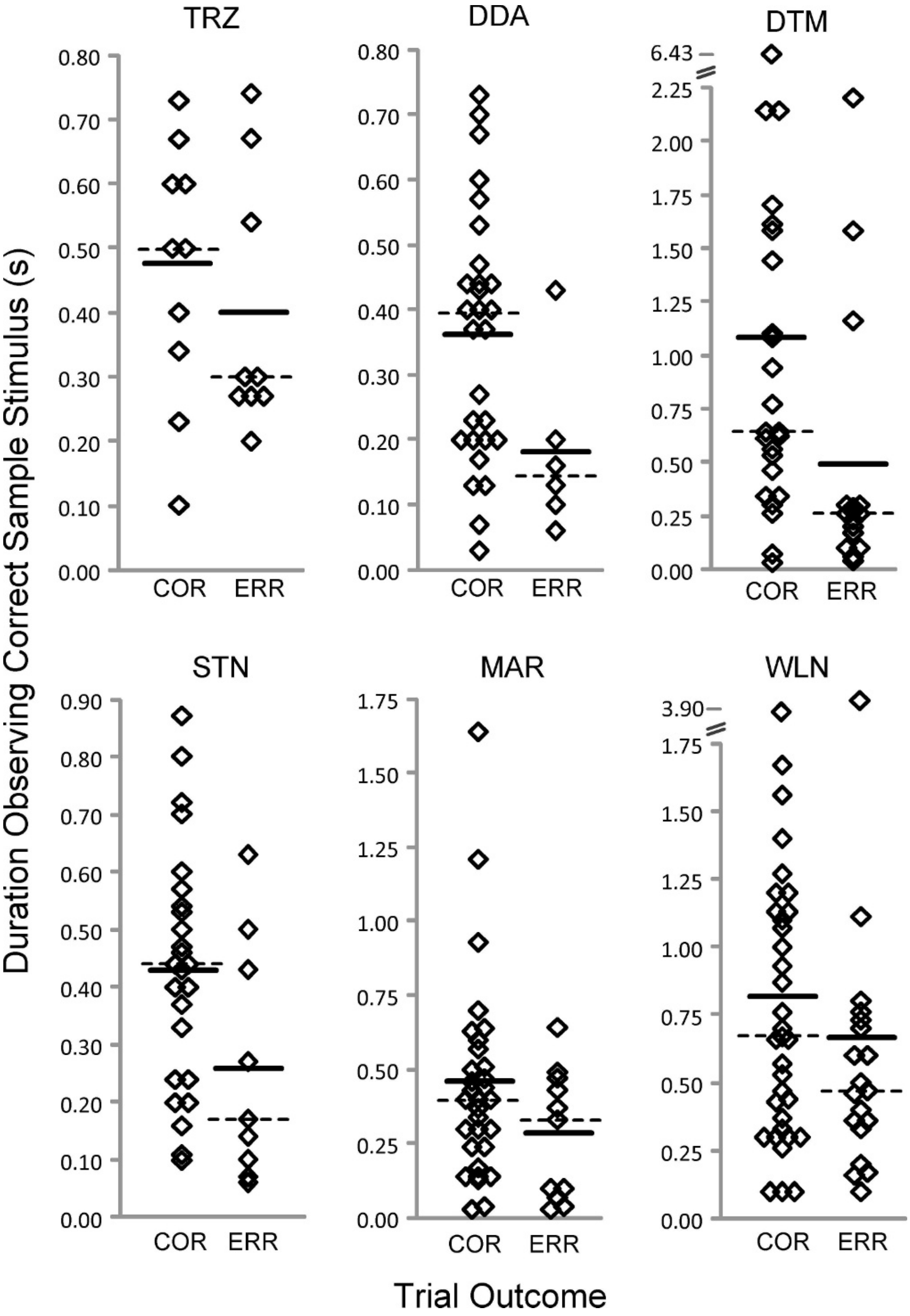

Fig. 2. Observing durations for sample stimuli that subsequently appeared as correct comparison stimuli. The distributions in the left and right portions of each panel show trials with correct and error responses, respectively. Solid and dashed horizontal lines within each distribution show mean and median durations, respectively. The data are from trials during the initial baseline sessions on which the participant observed both sample stimuli.

Table 5

Results of Intervention Procedures.

Participant and Condition	Number of Sessions	Mean Accuracy (%)	% Trials with One Observation	Observing Duration per Trial (s)	
				L	R
Participant MAR					
Baseline	5	68	37	0.48	0.21
Differential reinforcement for observing	9	63	35	0.36	0.25
Baseline	3	53	39	0.49	0.14
Extrastimulus prompts	1	88	0	2.41	1.51
Extrastimulus prompts with fading	8	83	0	1.32	1.33
Extrastimulus prompts	3	93	0	0.89	1.17
Baseline	3	67	38	0.71	0.20
Extrastimulus prompts	3	88	0	1.35	1.65
Baseline	4	70	35	0.82	0.16
Participant DDA					
Baseline	3	76	45	0.51	0.17
Within-stimulus prompts	1	97	0	2.80	1.73
Baseline	7	78	46	0.59	0.21
Within-stimulus prompts	4	94	0	3.54	1.90
Participant DTM					
Baseline	3	61	47	0.42	0.90
Within-stimulus prompts	2	78	4	1.66	2.63
Baseline	1	69	[a]		
Extrastimulus prompts	3	74	0	2.80	3.52
Within-stimulus prompts plus observing contingency	2	93	0	7.88	6.80
Baseline	3	77	68	0.75	0.23
Participant WLN					
Baseline	3	68	32	0.89	0.31
Within-stimulus prompts	3	85	1	5.44	1.56
Baseline	6	74	31	1.08	0.40
Observing contingency	2	83	0	3.98	2.61
Within-stimulus prompts plus observing contingency	3	94	0	6.79	4.24
Baseline	5	69	44	0.75	0.24
Participant STN					
Baseline	4	72	59	0.15	0.54
Within-stimulus prompts	4	91	0	4.37	3.28
Within-stimulus prompts, Fading Step 1	2	96	2	2.76	3.29
Within-stimulus prompts, Fading Step 2	3	77	13	1.37	2.36
Within-stimulus prompts (no fading)	3	64	0	1.86	3.67
Within-stimulus prompts, Fading Step 1 and High-accuracy contingency	4	93	3	2.14	3.19
Baseline	1	75	35	0.21	0.56
Within-stimulus prompts, Fading Step 2 and High-accuracy contingency	3	89	1	2.27	3.03

[a] Without eye tracking.

trials on which the participant failed to observe one of the sample stimuli, and the mean observing durations per trial for the stimuli in the left and right sample positions.

Participant MAR. The first intervention was differential reinforcement for observing, and over nine sessions there was little change from Baseline in accuracy scores, percent of trials with only one observation, or observing durations. After a return to Baseline for three sessions, again with little change in accuracy scores or observing behavior, MAR was given both vocal and gestural extrastimulus prompts. As Table 5 shows, in the first session observing durations increased to 2.41 and 1.51 s per trial for left and right sample stimuli, respectively, there were no failures to observe a sample stimulus, and accuracy improved to 88%.

During the eight sessions that followed, there were several attempts to fade the gestural prompts by gradually increasing the distance between the experimenter's hand and the sample stimuli (e.g., Schreibman, 1975). These attempts were not effective; MAR's gaze typically followed the experimenter's hand as it moved away from the sample stimuli. Mean accuracy during these sessions was 83%. The fading attempts were subsequently discontinued, and in the final three sessions with both vocal and gestural prompts on all trials accuracy was 93%, there were no observing failures, and observing durations were substantially longer than during Baseline. During a three-session return to Baseline, accuracy, rate of observing failures, and observing duration for the right-side sample stimulus reverted to previous levels. Subsequent prompt and Baseline conditions replicated the behavioral changes.

Participant DDA. The within-stimulus color prompt procedure was immediately effective. During sample display periods, DDA's point of gaze typically shifted back and forth to observe the darker sample stimulus. Accuracy and observing behavior improved immediately, as shown in Table 5. DDA then received a series of 27 Baseline sessions but eye tracking was conducted in only seven of these sessions because of technical difficulties with the apparatus; the results of these seven sessions are reported in Table 5 and they are similar to the initial Baseline condition. Accuracy scores were comparable in the 7 sessions with eye tracking (mean 78%, Table 5) and the 20 sessions without eye tracking (mean 79%, data not shown in Table 5). Table 5 shows that the within-stimulus color prompts were again effective when they were reintroduced.

Participant DTM. Table 5 shows that the within-stimulus color prompt procedure virtually eliminated DTM's observing failures and increased observing durations to the target range, but accuracy remained at intermediate levels (78%). Although DTM observed the sample stimuli, he typically did not follow the prompts and often looked away from the samples and at other items in the room (the tokens, the monitor housing, etc.); his mean duration of looking away from the samples was 2.46 s per trial in the first session with within-stimulus prompts and increased to 5.50 s per trial in the second session (looking-away durations are not shown in Table 5). Following a Baseline session, DTM received three sessions of extrastimulus prompts in the form of verbal instructions. These instructions were effective in that they controlled observing behavior. There were no observing failures, mean observing durations were 2.8 and 3.52 s per trial for the left and right sample positions (Table 5), respectively, and looking away decreased to 1.13 s per trial (not shown in Table 5). Apparently, DTM was observing but not attending consistently.

To encourage attending, the within-stimulus prompt procedure was modified to include the observing contingency for fixations of each sample stimulus during the prompt sequence. With the combined within-stimulus prompts and observing contingency, DTM's mean observing duration increased to 7.88 s and 6.8 s per trial for the left and right positions, respectively, and accuracy improved to 93%. In a final return to the Baseline condition, observing failures occurred on a majority of the trials and the increases in observing durations and accuracy scores were not maintained.

Participant WLN. Accuracy scores improved to 85% when within-stimulus color and size prompts were introduced. With a return to Baseline conditions, observing failures again occurred and accuracy declined to 74%. The Baseline procedure was then modified to include the observing contingency for two sessions. Observing failures were eliminated and durations increased, but accuracy improved only to 83%. He then received the combined within-stimulus prompts and observing contingency, and results were similar to those with DTM, a large increase in observing duration and high accuracy scores while the procedures were in effect, followed by a return to baseline observing behavior and accuracy when the procedures were withdrawn.

Participant STN was given the within-stimulus color and size prompt procedure first, and her initial response was very similar to Participant DDA's: large increases in observing durations, observing failures were completely eliminated, and accuracy increased to 91%. After four sessions with high accuracy, we attempted to fade out the within-stimulus prompts by gradual decreases in the size and color differences in the sample stimuli, as described in the Method section above. Accuracy remained high during the first fading step. With the second fading step, however, accuracy decreased to 77% and observing failures began to occur more often. Although a return to the full within-stimulus prompt procedure for three sessions eliminated the observing failures, accuracy remained low (64%). During one of these sessions, STN's session accuracy score was 28% and this included a run of 20 consecutive incorrect trials. This result suggested a motivational problem, and the high-accuracy contingency was initiated for this reason. STN, who exchanged her tokens for money after sessions, was told before each session that she would earn an extra dollar if her score was at least 33 points. Accuracy immediately recovered to high levels. Control by the intervention was replicated by a return to Baseline and reintroduction of the within-stimulus prompts with high-accuracy contingency. Thus, the initial results with STN were similar to those with Participants MAR and DDA in that contingencies that increased observing behavior also increased accuracy scores. These contingencies failed to sustain high accuracy in STN over subsequent sessions, however, but accuracy recovered with the implementation of a contingency that targeted accuracy specifically.

DISCUSSION

Initial Baseline evaluation. The results of the initial Baseline evaluation document a relation between two-sample DMTS scores indicative of atypically restricted stimulus control (stimulus overselectivity) and deficiencies in observing behavior. Analyses of eye movements showed that restricted stimulus control was accompanied by failures to observe all of the relevant stimuli and a tendency to observe for shorter durations than participants who did not exhibit restricted stimulus control. The analysis of incomplete observing in Table 4 shows that the overall intermediate accuracy scores resulted in part from averaging together trials on which the correct sample stimulus was observed and accuracy was very high, with trials on which the correct sample stimulus was not observed and accuracy was near chance levels. These data confirm previous interpretations of the relation between intermediate accuracy scores with the two-sample DMTS procedure and stimulus control (e.g., Dube & McIlvane, 1997).

The data on accuracy and observing durations shown in Table 2 and Figure 2 show a tendency for higher accuracy following longer observing durations. As noted above, interpretation of these data is complicated by the difference between the observed number of errors and the number of stimulus-control failures. All trials that end with errors provide unambiguous evidence of failures for sample stimuli to control responses to the comparison stimuli, but all trials that end with correct responses do not necessarily indicate stimulus control if the participant also makes a substantial number of errors. It seems reasonable to conclude only that the likelihood of stimulus control increases with longer observing durations. That is, as the duration of observing behavior increases (as measured by eye orientation toward stimuli) the probability of attending behavior increases (as measured by stimulus control by those stimuli). Clearly, however, there are other variables related to attending; Figure 2 shows that all participants made errors on a few trials in which the observing duration was within the upper half of the distribution for trials with correct responses.

These findings raise the question of whether undetected instances of incomplete observing and restricted stimulus control may be a source of the behavioral variability noted in individuals with developmental limitations responding to relational learning tasks (e.g., Dube & McIlvane, 1995). Suppose one were interested, for example, in establishing two equivalence classes, animals and food items, that included printed words. During the initial baseline training, trials would have to be arranged carefully to assure that stimulus control by words including common or similar letters (e.g., PIG, PIE, HEN, HAM) was not restricted to individual letters or letter combinations (see Sidman, 1971, for an early use of such procedures). Without such procedures, emergence tests presenting new combinations of baseline stimuli might be compromised by trials in which the controlling stimuli from different classes had overlapping features (e.g., requiring discrimination of PIG vs. PIE when the controlling features of both words during baseline training had been "P" or "PI"). This simple example is intended to illustrate a more general problem that may arise in relational learning tasks, including those that use abstract stimuli for experimental control purposes. While stimulus-feature overlap is obvious in the case of familiar printed letters, it may be difficult to detect with abstract stimuli that have common physical features. Nevertheless, the implications of restricted stimulus control in these situations would be similar to those in the simple example just presented.

Intervention procedures. The results of the intervention procedures showed that contingencies with clear effects on observing behavior most often produced comparable effects in stimulus control. The within- or extrastimulus prompting procedures immediately eliminated observing failures and produced observing durations that exceeded those of participants with high Baseline accuracy scores. In four of five instances, the introduction of these prompting procedures also improved accuracy scores substantially above baseline levels and to levels that indicate reliable stimulus control (DDA, 76% to 97%; STN, 72% to 91%; MAR, 53% to 88%; WLN, 68% to 85%). In these instances, the changes in observing behavior were accompanied by effective attending behavior. Thus, restricted stimulus control was shown to be correctable through training—at least temporarily—for these 4 individuals.

For Participant DTM, initial exposures to the within- and extrastimulus (verbal) prompting procedures had the predicted effect on observing but there was no meaningful change in accuracy. Results with DTM document apparent independence of observing and attending. (The qualifying ''apparent'' is used because it is possible that DTM did in fact attend to the stimuli he observed, but the comparison selection response that followed was controlled by some other unmeasured variable.) This result seems similar to those trials shown in Figure 2 on which observations for durations generally sufficient for stimulus control were nevertheless followed by errors.

For both DTM and WLN, the combination of the observing contingency and the within-stimulus prompt procedure produced observing durations that were much greater than necessary to meet the contingency, with left-plus-right totals of over 14 s per trial for DTM and 10 s for WLN. This effect may be due at least in part to changes in the stimulus control of the manual observing response—touching the sample stimulus—which terminated the sample display and thus the opportunity to observe the samples. The touchscreen was inactive during the prompting sequence, but when the procedure was introduced all participants touched the sample stimulus after their usual response latency. After some exposure to the procedure, DTM typically began to wait until the predictable 6-s sequence was completed and then touched. With the observing contingency, however, the prompting sequence was no longer regular and predictable. The sample display was sometimes static for several seconds at a time (if there was no fixation on the darker sample stimulus) and the animation-like effect of the regular sequence was attenuated. Visual cues as to when a sample touch would be effective were less salient, and DTM and WLN sometimes continued to observe the sample stimuli after the point in time at which a touch would have dismissed them. In situations such as this in which the participant controls the duration of stimulus display, this outcome suggests that remedial interventions for ineffective observing should take into account the possibility of independent stimulus control of the ocular observing behavior of interest and other behavior (e.g., manual) controlled by the same stimuli. That is, the behavior of observing the stimuli may be independent of the behavior of responding in other ways to the stimulus display (e.g., McIlvane, Kledaras, Callahan, & Dube, 2002).

To conclude, it is clear that at least some instances of restricted stimulus control are (a) due to incomplete and/or poorly organized observing behavior and (b) at least partially correctable through training. Interventions that control observing behavior seem necessary but may not always be sufficient for the remediation of restricted stimulus control in ID populations. Yet to be established is whether promising procedures of the type shown here and in other studies (e.g., Dube & McIlvane, 1999) can produce general, sustainable observing repertoires that lead to broadly effective observing and attending in the laboratory. If such an outcome can be achieved, it will set the stage for developing applied behavior analytic techniques that may have similar effects and perhaps profound functional significance in the world outside the laboratory.

REFERENCES

Allen, G., & Courchesne, E. (2003). Differential effects of developmental cerebellar abnormality on cognitive and motor functions in the cerebellum: An fMRI study of autism. *American Journal of Psychiatry, 160,* 262–273.

Barthold, C. H., & Egel, A. L. (2001). Stimulus overselectivity and generative language instruction for students with autism: An issue that needs to be revisited. *Behavior Analyst Today, 2*, 33–39.

Belmonte, M. K., & Yurgelun-Todd, D. A. (2003). Functional anatomy of impaired selective attention and compensatory processing in autism. *Cognitive Brain Research, 17*, 651–664.

Bickel, W. K., Richmond, G., Bell, J., & Brown, K. (1986). A microanalysis of the controlling stimulus–response relations engendered during the assessment of stimulus overselectivity. *The Psychological Record, 36*, 225–238.

Critchfield, T. S., & Perone, M. (1993). Verbal self-reports about matching to sample: Effects of the number of elements in a compound sample stimulus. *Journal of the Experimental Analysis of Behavior, 59*, 193–214.

Dickson, C. A., Wang, S. S., Lombard, K. M., & Dube, W. V. (2006). Overselective stimulus control in residential school students with intellectual disabilities. *Research in Developmental Disabilities, 27*, 618–631.

Dinsmoor, J. A. (1985). The role of observing and attention in establishing stimulus control. *Journal of the Experimental Analysis of Behavior, 43*, 365–381.

Dube, W. V. (2009). Stimulus overselectivity in discrimination learning. In P. Reed (Ed.), *Behavioral theories and interventions for autism* (pp. 23–46). New York: Nova Science Publishers.

Dube, W. V., Balsamo, L. M., Fowler, T. R., Dickson, C. A., Lombard, K. M., & Tomanari, G. Y. (2006). Observing behavior topography in delayed matching to multiple samples. *The Psychological Record, 56*, 233–244.

Dube, W. V., Iennaco, F. M., Rocco, F. J., Kledaras, J. B., & McIlvane, W. J. (1992). Microcomputer-based programmed instruction in identity matching to sample for persons with severe disabilities. *Journal of Behavioral Education, 2*, 29–51.

Dube, W. V., Lombard, K. M., Farren, K. M., Flusser, D. S., Balsamo, L. M., & Fowler, T. R. (1999). Eye tracking assessment of stimulus overselectivity in individuals with mental retardation. *Experimental Analysis of Human Behavior Bulletin, 17*, 8–14.

Dube, W. V., & McIlvane, W. J. (1995). Stimulus–reinforcer relations and emergent matching to sample. *The Psychological Record, 45*, 591–612.

Dube, W. V., & McIlvane, W. J. (1997). Reinforcer frequency and restricted stimulus control. *Journal of the Experimental Analysis of Behavior, 68*, 303–316.

Dube, W. V., & McIlvane, W. J. (1999). Reduction of stimulus overselectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis, 32*, 25–33.

Fetterman, J. G. (1991). Discrimination of temporal same–different relations by pigeons. In M. L. Commons, J. A. Nevin & M. C. Davison (Eds.), *Signal detection: Mechanisms, models, and applications* (pp. 79–101). Hillsdale, NJ: Erlbaum.

Howlin, P., Magiati, I., & Charman, T. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities, 114*, 23–41.

Lancioni, G. E., & Smeets, P. M. (1986). Procedures and parameters of errorless discrimination training with developmentally impaired individuals. In N. R. Ellis & N. W. Bray (Eds.), *International Review of Research in Mental Retardation, Vol. 14* (pp. 135–164). New York: Academic Press.

Litrownik, A. J., McInnis, E. T., Wetzel-Pritchard, A. M., & Filipelli, D. L. (1978). Restricted stimulus control and inferred attentional deficits in autistic and retarded children. *Journal of Abnormal Psychology, 87*, 554–562.

Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin, 86*, 1236–1254.

McHugh, L., & Reed, P. (2007). Age trends in stimulus overselectivity. *Journal of the Experimental Analysis of Behavior, 88*, 369–380.

McIlvane, W. J., Kledaras, J. B., Callahan, T. C., & Dube, W. V. (2002). High probability stimulus control topographies with delayed S+ onset in a simultaneous discrimination procedure. *Journal of the Experimental Analysis of Behavior, 77*, 189–198.

Reynolds, G. S. (1961). Attention in the pigeon. *Journal of the Experimental Analysis of Behavior, 4*, 203–208.

Schreibman, L. (1975). Effects of within-stimulus and extra-stimulus prompting on discrimination learning in autistic children. *Journal of Applied Behavior Analysis, 8*, 91–112.

Schreibman, L. (1997). The study of stimulus control in autism. In D. M. Baer & E. M. Pinkston (Eds.), *Environment and behavior* (pp. 203–209). Boulder, CO: Westview.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14*, 5–13.

Sidman, M., & Stoddard, L. T. (1966). Programming perception and learning for retarded children. In N. R. Ellis (Ed.), *International Review of Research in Mental Retardation, Vol. 2* (pp. 151–208). New York: Academic Press.

Theeuwes, J. (1994). Endogenous and exogenous control of visual selection. *Perception, 23*, 429–440.

Thorsteinsson, J. R., Martin, G. L., Yu, C. T., Spevack, S., Martin, T. L., & Lee, M. S. (2007). Predicting learning ability of people with intellectual disabilities: Assessment of Basic Learning Abilities test versus caregivers' predictions. *American Journal on Mental Retardation, 112*, 130–139.

Yantis, S. (1996). Attentional capture in vision. In A. F. Kramer, M. G. H. Coles & G. D. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 45–76). Washington, DC: APA.

Received: October 7, 2009
Final Acceptance: July 30, 2010