Published in final edited form as:

Am J Intellect Dev Disabil. 2016 May ; 121(3): 219–235. doi:10.1352/1944-7558-121.3.219.

STIMULUS OVERSELECTIVITY IN AUTISM, DOWN SYNDROME, AND TYPICAL DEVELOPMENT

William V. Dube, Rachel S. Farber, Marlana R. Mueller, Eileen Grant, Lucy Lorin, and Curtis K. Deutsch

University of Massachusetts Medical School – Shriver Center

Abstract

Stimulus overselectivity refers to maladaptive narrow attending that is a common learning problem among children with intellectual disabilities and frequently associated with autism. The present study contrasted overselectivity among groups of children with autism, Down syndrome, and typically developing children. Autism and Down syndrome groups were matched for intellectual level, and all three groups were matched for developmental levels on tests of nonverbal reasoning and receptive vocabulary. Delayed matching-to-sample tests presented color/form compounds, printed words, photographs of faces, Mayer-Johnson Picture Communication Symbols, and unfamiliar black forms. No significant differences among groups emerged for test accuracy scores. Overselectivity was not statistically overrepresented among individuals with autism in contrast to those with Down syndrome or typically developing children.

Keywords

stimulus overselectivity; attention; matching to sample; autism; Autism Spectrum Disorder; Down syndrome; children

Approximately 50 years ago, Dr. O. Ivar Lovaas and colleagues at UCLA were among the first to apply the methods of behavior analysis to the treatment and education of children with autism. The goals of their program were to establish communication in these children and to decrease their maladaptive behaviors (Smith & Eikeseth, 2011). One aspect of the children's behavior was described as follows:

Operationally, our data show that when autistic children are presented with multiple stimulus inputs, their behavior comes under the control of a range of input that is too restricted. This problem was referred to as “stimulus overselectivity” (Lovaas, Schreibman, Koegel, & Rhem, 1971) because the children overselected a limited number of stimuli from those available in their environment (Lovaas, Koegel, & Schreibman, 1979, p. 1237).

Since that time, the term *overselectivity* has frequently been associated particularly with autism. For example, parents of children with autism report overfocused patterns of

sensation and attention (Liss, Saulnier, Fein, & Kinsbourne, 2006). At one time overselectivity was proposed as the “keystone stimulus control deficit” in autism (Rincover, Feldman, & Eason, 1986). Further, overselectivity seems consistent with certain characteristics of Autistic Spectrum Disorders (ASD) such as highly restricted, fixated interests that are abnormal in intensity or focus (APA, 2013). Our study included individuals with Down syndrome (DS) as an intellectual disability contrast group, one that is often contrasted against ASD in the literature (Selzter, Abbeduto, Krauss, Greenburg, & Swe, 2004). Though attentional difficulties have been widely documented in DS, formal evaluation of overselective attending has been the focus of relatively little research.

Overselectivity may have an impact on several aspects of learning in special-education settings. The context for the present research study is the potential impact on discrete-trials instruction, a training method that is often utilized (among other teaching procedures) with students who have moderate to severe levels of intellectual disability. One problem is sometimes referred to as prompt dependency. When supplemental cues are given to prompt new learning (e.g., the teacher points to a referent while speaking the associated name), overselective attending only to the prompt (the teacher's pointing finger) and not to the referent may lead to errors when prompts are subsequently withdrawn (e.g., Schreibman, Charlop, & Koegel, 1982). A second potential problem is related to the generalization of learning. For example, if a student learns to identify an object by name on the basis of one isolated property of that object (e.g., a white eraser on a pencil), then the student may fail to identify other objects of the same class if they do not share that feature (e.g., pencils with different types of erasers). A related problem may arise with discriminations between items from different classes that share one feature, if the student has learned to identify one of them by that shared feature alone (examples in Dube & Wilkinson, 2014). Finally, the acquisition of language and communication skills may be impacted by overselectivity in discriminations of auditory (vocal speech) or visual (e.g., printed words) cues (e.g., Koegel, Schreibman, Britten, & Laitinen, 1979). For example, it may be discovered that the student who had apparently learned to recognize her printed name was actually responding to the initial letter only. To evaluate overselectivity in this context, test conditions were designed to model discrete-trials components of special-education curricula. The matching to sample procedure was selected in part because it is widely used in discrete-trials training to teach relations among words, pictures, and other symbols. The present study included some stimulus sets with academic relevance such as printed words and pictures.

Previous studies

Substantial previous research has shown that (a) overselectivity is much more likely in school-age or older children with autism than in typically developing children of the same chronological age (CA), and that (b) there is an inverse relation between mental age (MA) level and severity of overselectivity in children with intellectual disabilities both with and without autism (for recent reviews see Brown & Bebko, 2012; and Ploog, 2010). Relatively few research studies, however, were designed specifically to compare overselectivity in individuals with autism and intellectual disability versus intellectual disability without autism. These studies have employed three types of test procedures: free operant, simple discrimination, and conditional discrimination.

With free-operant procedures, stimuli are presented continuously for some fixed period of time, a continuously available response (e.g., button press) is intermittently reinforced, and the dependent measure is response rate. The origins of the idea that overselectivity occurs more often in autism than other intellectual disabilities may be traced to an early experiment using free operant procedures, Lovaas, Schreibman, Koegel, and Rehm (1971). Participants were six children with autism, five with intellectual disabilities including Down syndrome, and five typically developing children, all approximately the same CA (means 6–8 years). MA scores for the children with autism were not reported. The children were trained to press a lever in the presence of a multi-modal positive stimulus (S+) with auditory, visual, and tactile elements. Test trials measured relative response rates to the stimulus elements. Results showed that the typically developing children responded to all three elements, children with autism primarily to only one element, and the children with intellectual disability between these extremes. Greater overselectivity in autism was also reported using similar free-operant procedures in Frankel, Simmons, Fichter, and Freeman (1984). Participants were seven school-age children with autism (MA 3.5 years, IQ 68) and seven children with intellectual disability (mixed/unknown etiology, MA 3.8 years, IQ 60). The children with autism were more overselective than those with an intellectual disability. Matthews, Shute, and Rees (2001, Experiment 2) evaluated overselectivity in five adults with autism and five adults with intellectual disabilities (mixed/unknown etiology) matched for MA (mean 3.3 years). Stimuli were visual stimulus compounds that differed in size and location, test trials presented stimuli at intermediate sizes or locations, and results were expressed as generalization gradients. In contrast to other studies, the results showed less overselectivity in the adults with autism. The results with free-operant procedures are thus mixed, with some evidence for increased overselectivity in autism when stimuli are multi-modal but not within the visual stimulus dimension. Interpretation and general conclusions are complicated, however, by the relatively small numbers of subjects studied and differences among studies in subject characteristics.

Simple discrimination procedures have also been used to study overselectivity. In a typical experiment, multi-component positive (S+) and negative (S–) stimuli are presented on a series of training trials. After acquisition, stimulus control by the individual components, presented separately, is measured on test trials, typically without feedback (to prevent new learning during the test). Kovattana and Kraemer (1974) used such a procedure in a comparison of children with autism (no IQ or MA range reported), Down syndrome (IQ 30–53), and typically developing children. The procedure was a visual discrimination with three relevant stimulus dimensions of size, color, and form (e.g., large red circle vs. small blue triangle). After acquisition, test trials presented the stimulus dimensions separately (e.g., red vs. blue square to test color). Results showed that both discrimination learning and overselectivity in the children with autism was clearly dichotomous: Children with verbal abilities were not overselective and in fact equivalent to the typically developing children, but non-verbal children with autism required more training trials to learn the initial task and displayed more overselectivity than the children with Down syndrome. Because the study reported neither IQ nor MA scores for the non-verbal children with autism, however, those factors cannot be ruled out in accounting for the superiority of the Down syndrome group. Gersten (1983) used a similar visual simple-discrimination procedure to compare groups of

children with autism, intellectual disability of mixed/unknown etiology, and typically developing children. Groups were matched for CA (6–9 years) and expressive language. Both the autism and intellectual disability groups were more overselective than the typically developing children. There was, however, no significant difference between the autism and intellectual disability groups. Thus -- as is the case with the free-operant studies -- the results of studies using simple discriminations are mixed and difficult to interpret.

Conditional discrimination procedures: Matching to sample

Studies of overselectivity in individuals with intellectual disabilities using conditional-discrimination procedures such as matching to sample began to appear in the 1980s (e.g., Mackie & Mackay, 1982; Schneider & Salzberg, 1982; Whiteley, Zaparniuk, & Asmundson, 1987). In this type of procedure, multi-element or multidimensional sample stimuli are presented for an observation period, and then the sample is removed from view prior to presenting an array of comparison stimuli. On test trials, the comparison array is constructed so that either (a) only a subset of the sample elements or dimensions are displayed or (b) a subset of the sample elements or dimensions of interest are present in more than one comparison stimulus. Matching-to-sample procedures offer some significant advantages over simple-discrimination procedures, particularly for overselectivity evaluations in individuals with intellectual disabilities. Trials can be arranged so that all stimuli are correct and incorrect equally often and thus incorporate educationally relevant stimuli without the need to teach that specific stimuli are invariably incorrect (as with S– in simple discrimination; e.g., Dickson, Wang, Lombard, & Dube, 2006). Also, all correct responses on test trials can be followed by reinforcers, so that test contingencies are less likely to interact with motivation. As noted above, perhaps the greatest advantage of the matching-to-sample procedure is that it assesses overselectivity in a context that is highly relevant to the relational stimulus control of discrete-trials teaching procedures in special education.

Despite the advantages of matching to sample, only one study has used this method for a formal comparison of overselectivity in children with autism versus other intellectual disabilities. Litrowink, McInnis, Wetzel-Pritchard, and Filipelli (1978) compared overselectivity in groups of seven children (a) diagnosed with autism, (b) with Down syndrome and matched for MA (means 37–42 months) and IQ (means 45–46), and (c) typically developing children matched for MA. In preliminary training the children were taught to match identical vertical versus horizontal lines in a two-choice task. Test trials then presented stimuli that varied on four stimulus dimensions (color, shape, size, and number) in a 16-choice task. Both group and individual subject data analyses showed no significant differences between typically developing children and children with autism, but both of these groups showed significantly less overselectivity than the group of children with Down syndrome. One aspect of the procedure, however, the sudden increase in number of choice stimuli per trial from 2 to 16, introduces the possibility that the poorer performance in children with Down syndrome may be related to a delay or failure to generalize trained performances to the new format. Nevertheless, the study does raise the question of whether breadth of stimulus control in children with autism could actually be superior to that in children with Down syndrome in a relational learning context.

This possibility is inconsistent with the matching-to-sample results in Dickson, Wang, et al. (2006), in which overselectivity was found in 12 of 15 children meeting the clinical criterion for autism on the Autism Diagnostic Observation Schedule (Lord, Rutter, DiLavore, & Risi, 2001), but only in 6 of 14 not within the autism range. However, participants in this study were a typical cross-section of residential students, and those with versus without autism were not matched for MA. Thus, the results are merely suggestive of increased overselectivity in autism.

To summarize, the research literature has relatively few studies that directly compared overselectivity in individuals with autism versus other intellectual disabilities, and the results of these studies have been ambiguous and in some cases contradictory. In the domain of relational learning, which is most directly related to the discrimination requirements of special education curricula, there is no clear answer to the question of whether children with autism should be expected to exhibit a narrower range of stimulus control than children with other conditions associated with intellectual disabilities.

Rationale for present study

The present study used matching-to-sample test procedures to assess overselectivity in children with autism and compare the results to children with Down syndrome matched for both intellectual and developmental levels, and typically developing children matched for developmental level. Down syndrome was selected as a clinical contrast population because it is a well-defined neurodevelopmental disorder with a behavioral phenotype that has minimal overlap with autism. The autism versus Down syndrome contrast also allowed a re-examination of overselectivity in relational discriminations in these populations as in Litrowink et al. (1978), but in a testing context that verified behavioral prerequisites for the tests.

As noted earlier, the assessment was designed to model discrete-trials training with matching-to-sample procedures. For this reason, the ranges of intellectual disability and developmental levels for participants with autism and Down syndrome were selected to be representative of special-education students for whom discrete-trials teaching would be appropriate. Also for this reason, the assessment included tests with sets of stimuli relevant to discrete-trials training. Printed words were included because they are explicitly academic stimuli. Photos of faces and Mayer-Johnson Picture Communication Symbols (PCS; Mayer-Johnson, 2008) were included because they are often used in picture-based augmentative and alternative communication methods (e.g., Picture Exchange Communication System; Bondy & Frost, 2001), the former to represent specific people (peers, teachers, family, etc.) and the latter to represent a wide variety of objects, actions, and relations. Teaching special-education students to use such systems often involves discrete trials training. Finally, we also included tests with a set of abstract black-and-white forms generated for research purposes and unlikely to be familiar to participants; these were included to evaluate overselectivity with novel stimuli unlikely to have a learning or reinforcement history. Because overselectivity has been related to stimulus complexity (e.g., Dube et al., 2006; Ploog, 2010; Reed, Petrina, & McHugh, 2011), we predicted a range of accuracy scores related to the complexity of the stimuli: High scores with the least complex stimuli, which were single

stimuli with only two relevant features of color and shape; intermediate scores with single stimuli that had multiple relevant features (printed words, photos of faces), lower scores with arrays of two stimuli, and the lowest scores with arrays of three stimuli. The testing methods included (a) preliminary tests to verify generalized matching-to-sample performance, (b) verified accurate discrimination of experimental stimuli under conditions unlikely to provoke overselectivity, (c) positive feedback following all correct responses throughout testing, and (d) data analysis procedures that distinguished between overselective stimulus control by a subset of sample stimuli or stimulus features versus a general loss of stimulus control by sample stimuli.

Method

Participants

Participants were recruited from a variety of sources, among them referrals from special-education programs, advertising, and outreach via conferences and parent organizations. Parents or other caregivers completed a demographic checklist that included previous diagnoses. Participants were included in the Down syndrome group based on parental report of pediatric diagnosis. If autism was indicated on the checklist, the participant was evaluated with the appropriate module of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2001). The ADOS was administered by licensed clinical psychologist who was research reliable on the ADOS. Participants were included in the Autism group if scores met the diagnostic criteria for autism in the communication domain, the social interaction domain, and for the combined communication plus social interaction total. Participants with Down syndrome and typically developing children were screened with the Childhood Autism Rating Scale - 2 (CARS; Schopler, Van Bourgondien, Wellman, & Love, 2010) and included in those groups if the results were below the cutoff score for the diagnosis of autism.

Intellectual level was measured by Differential Abilities Scales II (DAS-II; Elliott, 2007) General Conceptual Ability standardized composite scores (GCA; mean = 100, SD = 15; comparable DAS-II Special Nonverbal Composite scores were substituted for eight nonverbal participants with autism and one with Down syndrome). Scores for participants 18 years or older were calculated using the normative data for age 17 years, 11 months. Participants with intellectual disabilities were included if GCA scores were within the range of 30 to 65. This range, which is more than two standard deviations below the mean, was selected because special-education students with this level of intellectual disability are the most likely candidates for discrete-trials instruction. Typically developing children were included if scores were at least 85 (one standard deviation below the mean).

Developmental level was characterized by (a) receptive vocabulary via the Peabody Picture Vocabulary Test 4 (PPVT; Dunn & Dunn, 2007) Growth Scale Value scores (GSV), and (b) nonverbal reasoning via the DAS-II Matrices subtest ability scores. The PPVT and DAS-II Matrices subtest were chosen in part because both tests could be administered to participants with and without intellectual disabilities (the Matrices subtest is included in both the Early Years and School Age versions of the DAS-II). The PPVT was administered immediately after the initial screening (see below). The CARS, ADOS, and DAS-II tests were

administered after completion of the discrimination pretests (also described below); the order of ADOS and DAS-II testing was uncontrolled.

Setting, Apparatus, and Test Sessions

Testing was conducted in quiet areas containing a table and two chairs at the participant's school, at a research center, or in the home. The test apparatus was portable and transported to the testing site in a rolling suitcase. Participants sat at the table in front of a 15" LCD color touch screen monitor (Elo 1515L). Test procedures were presented by a laptop computer running custom software written in MATLAB (MathWorks, Inc.) and the Psychophysics Toolbox (Brainard, 1997). The overall duration of test sessions was 20–30 minutes and each session typically included three to five blocks of matching-to-sample trials (described below).

Procedures

This section describes the matching-to-sample test procedure and consequences for correct responses, an initial screening procedure that introduced the test situation and the matching task, discrimination pretests with stimuli that were similar to those to be used in the overselectivity assessments, an overview of the assessment strategy, and finally detailed descriptions of the assessment tests.

Matching to sample procedure—The test procedure was three-choice identity matching to sample. After initial screening (described below), sessions consisted of one or more blocks of matching-to-sample trials, with 18–40 trials per block. Trials began when a sample stimulus was presented at the top of the screen. There was no limit on sample stimulus duration; the sample remained displayed until the participant touched it. When the participant touched the sample: (a) in simultaneous matching to sample (SMTS), the sample remained displayed and three comparison stimuli were presented in a horizontal array at the bottom of the screen; (b) in delayed matching to sample (DMTS), the sample disappeared and the three comparison stimuli were presented immediately at the bottom of the screen (i.e., 0-s delay). On all trials throughout the experiment, the correct comparison stimulus (S+) was completely identical to the sample and neither of the two incorrect comparison stimuli (S–) was completely identical. The correct response was touching the S+ comparison stimulus. Within each block of trials, the order of sample stimuli and S+ comparison locations was unsystematic with the restrictions that a specific sample stimulus or S+ location was repeated on no more than three consecutive trials. Comparison stimuli appeared in each location equally often and in an unsystematic order, with the restriction that each comparison position was correct equally often.

Consequences—Prior to participation, caregivers were asked to suggest several appropriate items likely to serve as reinforcers for each individual participant, including snack foods, trinkets, and activities (e.g., playing a computer game). Before each session, participants were shown the available items. During all sessions (including initial screening, pretest, Baseline, and Test phases), 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 stimulus presentation. Incorrect responses were followed by a 2-s black screen and then the inter-trial interval. Accumulated tokens were exchanged for the participant's choice(s) among the available items after each block of trials.

Initial screening—All participants were screened to determine whether they were able to perform identity matching to sample and to introduce the tokens and token exchange procedure. In language appropriate for the participant, the experimenter explained that the he or she could earn tokens for playing a matching game on the computer and exchange them later for the available items. Screening stimuli were 6×8 cm (vertical $\times$ horizontal) color pictures of a dog, turtle, and fish. SMTS was introduced first. When the sample for the first trial appeared, the experimenter said, “Look at this picture and then touch it. Now look at these three pictures and touch the one that's the same.” Verbal, gestural, or physical prompts were included as needed. A correct response was followed by praise and presentation of a token. If the participant touched the monitor screen during the inter-trial interval, the experimenter gave prompts to place hands down on the table. Over successive trials, the praise was eliminated and any prompts were gradually withdrawn. If the participant completed three consecutive trials correctly without prompting, then screening continued with the DMTS procedure. Each session included 6 to 18 trials and the criterion for continuing in the experiment was three consecutive correct DMTS trials without prompting, within a limit of three screening sessions.

Discrimination pretests were conducted to verify that participants were willing and able to perform the identity matching-to-sample task with stimuli that were similar to (but different from) those to be used in the overselectivity assessments. The procedure was DMTS. Stimuli for the Icon Pretest were 4.5×5.5 cm color Mayer-Johnson Picture Communication Symbols (PCS) for flowers, chair, and brush (Mayer-Johnson, 2008); for the Color Pretest, 4.5×5.5 cm rectangles filled with red, blue, green, gray, brown, or yellow; for the Printed Word Pretest, the words CAB, TOP, and MEN in black 48-point Arial font (1.6 cm in height). The pretests were presented in blocks of 18 trials. The criterion to pass a pretest was one block with 16 or more correct responses, within a testing limit of three blocks. This limit was extended if accuracy was increasing across successive blocks.

Overview of assessment strategy—After the initial screening and pretests, overselectivity was assessed in five test series with five different sets of stimuli. Each test series included Baseline and Test phases. High Baseline accuracy was required to continue to the Test phase.

Three of the test series evaluated overselectivity with multiple features within stimuli. The multiple-feature stimulus sets included (a) color/form compounds, (b) three-letter printed words, and (c) photos of faces. During the Baseline phases, the comparison stimuli did not share any of the features to be tested. During the Test phases, one of the incorrect (i.e., non-matching) comparison stimuli was altered so that some but not all of the features were identical to the sample and correct comparison stimulus. Thus, overselective attending to the features shared by a correct and incorrect comparison would produce errors on some trials.

Two additional test series assessed overselectivity with arrays of two or three stimuli: (d) PCS symbols used in augmentative and alternative communication and (e) unfamiliar black-and-white forms. During Baseline phases, one sample stimulus was presented on each trial, and the comparisons consisted of three individual stimuli. During Test phases, samples consisted of arrays of two or three stimuli, only one of which was present in the subsequent comparison array. Thus, overselective attending to only a subset of the stimuli in the sample arrays would produce errors on some trials.

Color/form Compounds test—Prerequisites were the Icon and Color Pretests. Baseline stimuli were six PCS symbols, approximately 4.5×5.5 cm, and edited to consist of black lines and one other color per symbol: blue ball, brown book, gray cup, green lamp, red stapler, and yellow telephone. The Baseline procedure was DMTS and the accuracy criterion to continue to the Test phase was at least 89% (16/18) correct within three 18-trial blocks. For the DMTS Test phase, each trial included one unaltered S– baseline comparison and one S– comparison altered to be the identical color or shape as the S+. For example, on one of the color test trials with the blue ball as sample, the comparisons were a blue ball (S+), blue book (altered S–), and gray cup (unaltered baseline S–); on one of the shape test trials with the blue ball as sample, the comparisons were a blue ball (S+), brown ball (altered S–), and gray cup (unaltered baseline S–). Each of the 60 possible test trials was presented one time, in two blocks of 30 trials each. Each block included 15 color test trials and 15 shape test trials.

Printed Word test—Prerequisites were the Icon and Printed Word Pretests. Baseline stimuli were SAN, DEH, FOL, CIV, and BUR displayed in black 48-point Arial font. The Baseline procedures were SMTS, followed by DMTS, and the accuracy criterion to continue was at least 87% (13/15) within three 15-trial blocks with each procedure. For the DMTS Test phase, each trial included one unaltered S– baseline comparison and one S– comparison altered to have the identical initial letter or final two letters as the S+. For example, on one of the initial-letter test trials with SAN as sample, the comparisons were SAN (S+), SEH (altered S–), and FOL (unaltered baseline S–); on one of the final-letters test trials with SAN as sample, the comparisons were SAN (S+), DAN (altered S–), and FOL (unaltered baseline S–). Each of the 40 possible test trials was presented twice, in two blocks of 40 trials each. Each block included 20 initial-letter test trials and 20 final-letters test trials.

Faces test—The Icon Pretest was the only prerequisite for administration of this test. Baseline stimuli were three 9.5×6.5 cm black-and-white photos of female faces, shown in Figure 1 (from Le Grande, Mondloch, Maurer, & Brent, 2003). The Baseline procedures were SMTS, followed by DMTS, and the accuracy criterion to continue was at least 89% within three 18-trial blocks with each procedure. For the DMTS Test phase, each trial included one unaltered S– baseline comparison and one S– comparison altered to have the identical eyes portion or mouth portion as the S+. For example, on one of the eyes test trials with Face A as sample, the comparisons were Face A (S+), Face B with the eyes of Face A (altered S–; see Fig. 1), and Face C (unaltered baseline S–); on one of the mouth test trials with Face A as sample, the comparisons were Face A (S+), Face B with the mouth of Face A

(altered S–), and Face C (unaltered baseline S–). Each of the 12 possible test trials was presented six times, in two blocks of 36 trials each. Each block included 18 eyes test trials and 18 mouth test trials.

Color Symbols test—Prerequisites were the Icon and Color Pretests. Stimuli were 180 PCS symbols approximately 2.5×2.5 cm. Baseline and Test phases used trial-unique stimulus presentation; the stimuli for each block of trials were drawn at random without replacement from the pool of 180 symbols, and different stimuli were presented on each trial. The Baseline procedure was DMTS with one sample stimulus per trial, and the accuracy criterion to continue was at least 89% within three 18-trial blocks. For the DMTS Test phase, each trial displayed two or three sample stimuli in a horizontal array, with approximately 2.5 cm between adjacent stimuli. On each trial, only one of the sample stimuli appeared in the comparison array as S+, and the two S– stimuli had not been displayed previously within the trial block. The S+ was identical to the sample stimulus in each sample position (left, center, right) equally often. The Test consisted of three blocks of 36 trials (108 trials total) and each block included 18 two-sample trials and 18 three-sample trials, interspersed and alternating irregularly.

Unfamiliar Forms test—The Icon Pretest was the only prerequisite. Stimuli were 180 black forms approximately 2×2 cm presumed to be unfamiliar to participants; examples are shown in Figure 1 (also see Dube & McIlvane, 1999; Dickson, Deutsch, et al., 2006). During the Test phase the sample stimuli were two or three forms presented in a horizontal array with approximately 2 cm between adjacent forms. All other details of the procedure were the same as the Color Symbols test.

Lost Stimulus Control Criteria

High accuracy scores on the overselectivity tests indicate attending to all of the sample stimuli or relevant stimulus features, and thus no overselectivity. The test procedures were designed so that attending to some but not all of the stimuli or features would produce intermediate accuracy scores indicative of overselectivity. Very low accuracy scores, however, do not provide evidence for overselective stimulus control because they indicate a general loss of stimulus control by the experimental stimuli. Previous studies have addressed this issue. For example, the data analyses in both Reed, Stahmer, Suhrheinrich, & Schreibman (2013) and Rieth, Stahmer, Suhrheinrich, & Schreibman (2015), with simple-discrimination procedures, incorporated criteria to omit data sets for participants who failed to maintain baseline discrimination accuracy during probe tests for overselectivity. This section describes *lost stimulus control criteria* for omitting uninterpretable data with matching-to-sample procedures. For each test, participants were added as needed to replace data sets that were omitted based on these criteria.

For the multiple-feature tests (Color/form Compounds, Printed Word, Faces), the comparison arrays during the Test phases included one stimulus that was identical to the sample (S+), one S– that was altered to include some but not all features of the sample, and one unaltered baseline S– stimulus. The criterion for lost stimulus control included two components: (a) test accuracy was less than 85% and (b) more than 30% of the errors were

selections of unaltered S– baseline stimuli. High accuracy during the initial Baseline phase verified not only reliable selection of the S+ comparison, but also reliable rejection of the unaltered non-matching S– comparisons. The emergence of a substantial proportion of errors made by selecting unaltered stimuli during the Test phase was interpreted as a general loss of stimulus control by sample stimuli; in such cases errors made by selecting the altered S– stimulus become uninterpretable.

For the tests with two-sample arrays, overselective attending to only one sample on every trial would result in (a) approximately 100% correct responding when that stimulus appeared as the S+ comparison, but (b) chance level of 33% when that stimulus did not appear in the comparison stimulus array. With both types of trials equally represented, the overall accuracy score would be approximately 67% (see Dickson, Deutsch, et al., 2006 for further details of the accuracy score analysis). For the three-sample arrays, overselective attending to only one sample on every trial would result in an overall accuracy score of approximately 55%. Substantially lower accuracy scores that approach chance level (33%), however, indicate a general loss of stimulus control by sample stimuli, and thus they cannot be interpreted as overselective stimulus control by those stimuli. The lost stimulus control criterion for accuracy that approached chance levels was set at 50% for the two-sample tests and 40% for the three-sample tests. These cutoff scores are approximately halfway between the score that would result from stimulus control by only one sample on every trial (as above) and the score that would result from pure chance (i.e., no control by sample stimuli on any of the trials).

Results

Initial Screening and Discrimination Pretests

Initial screening and preliminary testing was conducted with 109 potential participants, including 78 individuals age 9–21 years with intellectual disabilities and 31 typically developing children age 3.5–8 years. During initial screening, 11 potential participants with intellectual disability were withdrawn for behavioral reasons such as disruptive or interfering behavior (refusing to remain seated, persistent inability or unwillingness to point to pictures on the touch screen, etc.). Another five potential participants with intellectual disability were withdrawn during initial screening because they did not meet the accuracy criterion.

High accuracy on the Icon Pretest was a prerequisite for all of the overselectivity tests, and 10 potential participants with intellectual disability were withdrawn because they failed this pretest. One participant with intellectual disability and three typically developing children failed the Printed Word Pretest; they continued in the study but they were not given the Printed Word test. Thus, a total of 83 participants completed the initial screening and preliminary testing and received at least one of the experimental tests; most of these participants received more than one test. The top row of Table 1 shows that this pool of participants included 29 for experimental groups designated Autism, 23 for DS (Down syndrome), and 31 for TDC (typically developing children). For each overselectivity test, Table 1 also shows the numbers of participants within each group who started the Baseline phase, failed to meet accuracy criteria for Baseline phases, and the number of data sets omitted from test analyses due to lost stimulus control criteria. When Baseline failure or lost

stimulus control occurred, the participant was withdrawn from the test and Baseline training was initiated with an additional participant.

Participant Characteristics for Overselectivity Tests

Table 2 summarizes descriptive statistics for subject demographics and matched variables for the three experimental groups (Autism, DS, and TDC) for each test. Group contrasts were performed for gender using a chi-square test across all groups. A Student t-test was used to contrast CA and DAS-II GCA scores for Autism vs. DS groups only. One-way analyses of variance (ANOVA) at three levels were used to contrast PPVT-4 GVS and DAS-II Matrices ability scores across all groups. Significance (p) levels for group differences are summarized for each of these dependent measures. For the Printed Words, Faces, and Unfamiliar Forms tests, there were significant differences in CA between Autism and DS groups, with Autism groups older. The only other significant difference was between Autism and DS groups for the Color Symbols test, with DAS-II CGA score higher for Autism, although scores for both groups were below three standard deviations from the mean. With this one exception, Table 2 shows that group matching was successful for intellectual disability and developmental level.

Results of Overselectivity Tests

Figure 2 summarizes the results of overselectivity Test phases as group mean accuracy scores for each test. Because errors on test trials indicate overselective stimulus control, high accuracy scores are interpreted as evidence for no overselectivity, intermediate scores in the range of 75–85% as relatively mild or infrequent overselectivity, and scores below 75% as evidence for a substantial impact of overselectivity on matching-to-sample performance. Color/form Compounds test accuracy scores were very high for all groups (> 95%). The Printed Words accuracy scores were high for Autism and DS groups (91.7% and 91.5%, respectively), and approached intermediate for TDC (86.1%). On the Faces test, accuracy scores were at the high end of the intermediate range and virtually identical for all groups (83.6% to 85.9%). Results with arrays of two Color Symbols were also intermediate for all groups, and those with three Color Symbols and two or three Unfamiliar Forms were relatively low. Thus, the overall results confirm the predicted relation between accuracy scores and stimulus complexity with the greatest impact on tests with multiple stimuli.

Table 3 shows accuracy scores listed by dependent measure along with inferential statistics for group contrasts. ANOVAs were used for contrasts, except in those instances in which Shapiro-Wilk tests indicated a deviation from a Gaussian distribution (Hays, 2007); in these cases, a Kruskal-Wallis test was used. Significance (p) levels for omnibus group tests are reported across experimental groups, followed by pairwise post-hoc contrasts for these groups. The latter include both the p-value and the Effect Size (d) for each contrast computed using the methods of Cohen (1988). For these analyses, significance values were adjusted to control for multiplicity using the Bonferroni correction; for each adjusted p-value in Table 3 the corresponding unadjusted significance level is listed in parentheses.

Table 3 shows that there were no significant differences in accuracy scores across diagnostic groups (Group Test column) or in any of the post-hoc pairwise comparisons. Despite the fact

that the differences were not statistically significant, medium to large effect sizes ($> .60$) were obtained on the Color/form Compounds and Printed Words tests. For both of these tests, accuracy scores were lower for TDC than for either of the groups with intellectual disability.

A recent review of the research literature on face identity recognition in ASD “found evidence that people with ASD have specific deficits discriminating eyes” (Weigelt, Koldewyn, & Kanwisher, 2012, p. 1081). To examine this possibility we conducted an additional analysis of mouth vs. eyes errors on the Faces test. Results showed that error distributions were the same in DS and TDC groups (22% baseline, 31% eyes, 47% mouth). The Autism group had a slightly larger proportion of eyes errors (43%) and smaller proportion of mouth errors (40%) than the other groups, though a Kruskal-Wallis test showed that these differences were not statistically significant, $\chi^2(2) = 1.44$, $p = 0.49$.

Discussion

Stimulus overselectivity was not more severe in the Autism groups than DS or TDC groups when (a) Autism and DS groups were matched for intellectual level (with the exception of one test), (b) all groups were matched on developmental measures of receptive vocabulary and nonverbal reasoning, (c) Baseline phase results verified discrimination of all experimental stimuli under conditions unlikely to provoke overselectivity, and (d) data analytic procedures distinguished between overselective stimulus control and a general failure of stimulus control. Our results support a hypothesis that, when the intellectual disability groups are matched as described above, levels of stimulus overselectivity appear to be comparable in pairwise contrasts, with no evidence of a statistical overrepresentation of overselective attending in autism. It seems worth noting that a recent reexamination of overselectivity in children with ASD -- also using data analytic procedures that distinguished between overselective stimulus control and lost stimulus control -- suggests that overselectivity may not be as prevalent in ASD as previously estimated (Rieth et al., 2015).

Although the accuracy scores within each test were similar across the three groups, accuracy varied widely across the test series (Fig. 2), ranging from very high accuracy on the Color/form Compounds test ($> 95\%$), to low accuracy on the Unfamiliar Forms tests with three sample stimuli ($< 60\%$). Taken together, the data from the test series support an interpretation that the degree of stimulus overselectivity in this study was related to stimulus complexity in terms of the number of relevant stimulus features and the number of stimuli, and that familiarity or experience with the stimulus types may also be a factor.

The present results are partially consistent with a previous study of overselectivity using matching-to-sample test procedures (Litrowink et al., 1978). As in the previous study, we found no differences between groups with autism and typically developing children matched for developmental level. In contrast to Litrowink et al., we found no differences between groups with autism and Down syndrome, whereas Litrowink et al. reported significantly greater overselectivity in their Down syndrome group. Possible reasons for the different findings may include small group sizes in Litrowink et al. ($N = 7$) but perhaps more importantly better adaptation to the testing procedures in the present study. As noted in the

introduction, the test procedure in Litrowink et al. involved a sudden increase in number of choice stimuli per trial from 2 to 16 and the poorer performance in children with Down syndrome may be related to poorer generalization of trained performances to a new test format rather than specifically to overselectivity. In the present study, all participants were familiarized with the test format prior to testing.

Autism and DS group matching on intellectual level failed for the Color Symbols test DAS-II GCA standardized scores, although the scores for both groups were more than three standard deviations below the mean (47.1 and 37.8, respectively) and thus indicate moderate to severe levels of intellectual disability for both groups. Matching was successful for this test on the developmental measures of PPVT-4 GSV and DAS-II Matrices ability scores. Although the differences between Autism and DS accuracy scores were not significant, Figure 2 and Table 3 show that DS accuracy scores were slightly lower on both two- and three-sample Color Symbol tests (differences < 5%) and the possibility of a relation between accuracy and intellectual level should be acknowledged. On the Unfamiliar Forms tests, both DAS-II GCA scores and accuracy scores were similarly slightly lower for DS than Autism, although neither of these differences was significant.

Although none of the group accuracy score differences reached statistical significance, pairwise contrasts showed moderate-to-large effect sizes for the Color/form Compounds and the Printed Words tests, with accuracy lower for TDC on both tests. Regarding the former, the accuracy scores were so high (95.1% to 97.7%) that the clinical significance of the difference is negligible. For the Printed Words test, accuracy scores were high for the Autism (91.7%) and DS (91.5%) groups, but the TDC group accuracy scores of 86.1% approached the intermediate range and thus may be interpreted as showing some evidence of overselectivity. One possible explanation is that the higher accuracy scores for the groups with intellectual disability were due in part to years of experience with similar stimuli in special-education settings; the 5-year-old typically developing children lacked equivalent experience.

The Color/form Compounds test results with typically developing children are consistent with a recent study by Reed et al. (2013). With simple-discrimination procedures and three-dimensional color/form compounds (wooden blocks), overselectivity occurred primarily in typically developing children younger than 36 months (the oldest with an overselective outcome was 38 months). The present study used somewhat more complex matching-to-sample procedures and two-dimensional color/form compounds, and there was no evidence of overselectivity even in the youngest children tested; mean accuracy score for the six youngest participants in the present study (range, 42–46 months) was 95% (range, 90–100%). Additional research with younger children will be needed to describe the developmental trajectory of overselectivity with two- versus three-dimensional color/form compounds and simple versus conditional discrimination procedures.

The present results with Unfamiliar Forms and three samples are consistent with those of Reed (2012) with similar stimuli (3-element arrays of Wingdings font characters, etc.) and similar procedures (delayed matching to sample). Mean accuracy score in Reed for 0-s delayed matching was approximately 76% for children with autism but no intellectual

disability (CA = 9 years, MA = 8.9 years), predictably higher than 57.1% in the present study for children with autism and moderate to severe intellectual disability. Also, the mean accuracy score in Reed on the same test was approximately 94% for typically developing 8-year-old children, again predictably higher than 58.8% for 4.7-year-old typically developing children in the present study.

The Weigelt et al. (2012) literature review on face identity recognition in ASD reported that the majority of studies of simple face perception found no autism deficit in tests with no memory demand, but deficits were consistently reported with even a low level of memory demand. Although the results of the present study may at first seem inconsistent with this literature, several points merit consideration. In the present study, the delay between sample stimulus offset and presentation of comparison stimuli was zero seconds and thus memory demand was at an absolute minimum. In order to qualify for the overselectivity test, participants first had to achieve high accuracy on the 0-s delayed matching task in the Baseline phase (in which all of the face photos in the comparison array had different eyes and mouths). Thus, the Baseline phase eliminated potential participants with poor face discrimination under conditions of minimal memory demand, as intended, so that the Test phase would measure overselective discrimination and not poor discrimination per se. Table 1 shows that Baseline failure rates were higher for the Faces test than any other test, but the failure rate for Autism (7/21, 33.3%) was no higher than the failure rates for DS (8/23, 34.8%) or TDC (10/23, 43.5%). The analysis of mouth vs. eyes errors showed a slightly larger proportion of eyes errors in the Autism group than the other groups (43% vs. 31%) but this difference was not significant. Because the test results showed no significant differences among groups in accuracy scores or error distribution, they provide no evidence for a deficit specific to autism. We also note that the Faces test was included to assess overselectivity as a pattern-recognition issue in the use of photos as symbols for people, and not intended to model overselectivity as it may affect social communication within interpersonal interactions.

This study was conducted with matching-to-sample procedures similar to those used in discrete-trials teaching procedures. Such procedures are components of many special-education curricula and used to teach relations among pictures, printed words, and symbols to students with levels of intellectual disabilities similar to those of the study samples (i.e., GCA or IQ scores in the 40–50 range). The implications for special education with discrete-trials procedures may be summarized as follows: First, learning problems related to overselectivity may occur with comparable frequency in students with intellectual disability with or without autism. When analyzing learning problems, the teacher may benefit from giving equal priority to the potential for overselectivity, regardless of autism diagnosis. Second, overselectivity may occur with some types of stimuli and not others, depending on the relationships among stimulus complexity, learning history, and developmental level.

Acknowledgments

Research and manuscript preparation was supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development under award numbers R01HD062582, P01HD025995, 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.

We thank Dr. Anne Hunt for statistical analyses through the UMMS Shriver Center's IDRC Quantitative Methods Core, Dr. Daphne Maurer for permission to use the stimuli for the Faces test, and Dr. Bill McIlvane for comments on the manuscript. Participant characterization was conducted by the Shriver Center's IDRC Clinical and Translational Research Support Core.

We thank the participants who met with us individually, as well as the staff and students of The New England Center for Children, Southborough, MA; The League School, Walpole, MA; The Evergreen Center, Milford, MA; May Institute, Randolph, MA; Children's Center for Communication Programs, Beverly, MA; Guild for Human Services, Waltham, MA; Crotched Mountain School, Manchester, NH; Mercy Centre, Worcester, MA; St. Coletta Day School, Braintree, MA; Archway, Leicester, MA; and Fletcher Maynard Academy, Cambridge Public Schools, Cambridge, MA.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed.. Author; Washington, DC: 2013.
- Bondy A, Frost L. The Picture Exchange Communication System. Behavior Modification. 2001; 25:725–744.
- Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]
- Brown SM, Bebko JM. Generalization, overselectivity, and discrimination in the autism phenotype: A review. Research in Autism Spectrum Disorders. 2012; 6:733–740.
- Cohen, J. Statistical power analysis for the behavioral sciences. 2nd ed.. Lawrence Erlbaum Associates; New York: 1988.
- 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. [PubMed: 17029502]
- Dickson CA, Wang SS, Lombard KM, Dube WV. Overselective stimulus control in residential school students with intellectual disabilities. Research in Developmental Disabilities. 2006; 27:618–631. [PubMed: 16290082]
- Dube WV, Balsamo LM, Fowler TR, Dickson CA, Lombard KM, Tomanari GY. Observing behavior topography in delayed matching to multiple samples. The Psychological Record. 2006; 56:233–244.
- Dube WV, McIlvane WJ. Reduction of stimulus overselectivity with nonverbal differential observing responses. Journal of Applied Behavior Analysis. 1999; 32:25–33. [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. [PubMed: 24773053]
- Dunn, LM.; Dunn, DM. *Peabody Picture Vocabulary Test*. 4th ed.. NCS Pearson; Minneapolis, MN: 2007.
- Elliott, CD. Differential Ability Scales. 2nd ed.. Harcourt Assessment; San Antonio, TX: 2007.
- Frankel F, Simmons J,Q, Fichter M, Freeman BJ. Stimulus overselectivity in autistic and mentally retarded children - a research note. Journal of Child Psychology and Psychiatry. 1984; 25:147–155. [PubMed: 6693522]
- Gersten RJ. Stimulus overselectivity in autistic, trainable mentally retarded, and non-handicapped children: Comparative research controlling chronological (rather than mental) age. Journal of Abnormal Child Psychology. 1983; 11:61–76. [PubMed: 6853882]
- Hays, W. Statistics. 6th ed.. Wadsworth Publishing; Independence, KY: 2007.
- Koegel RL, Schreibman L, Britten K, Laitinen R. The effects of schedule of reinforcement on stimulus overselectivity in autistic children. Journal of Autism and Developmental Disabilities. 1979; 9:383–396.
- Kovattana PM, Kraemer HC. Response to multiple visual cues of color, size, and form by autistic children. Journal of Autism and Childhood Schizophrenia. 1974; 4:251–261. [PubMed: 4280396]
- Le Grande R, Mondloch CJ, Maurer D, Brent HP. Expert face processing requires visual input to the right hemisphere during infancy. Nature Neuroscience. 2003; 6:1108–1112. [PubMed: 12958600]
- Liss M, Saulnier C, Fein D, Kinsbourne M. Sensory and attention abnormalities in autistic spectrum disorders. Autism. 2006; 10:155–172. [PubMed: 16613865]

- Litrownik AJ, McInnis ET, Wetzel-Pritchard AM, Filipelli DL. Restricted stimulus control and inferred attentional deficits in autistic and retarded children. *Journal of Abnormal Psychology*. 1978; 87:554–562. [PubMed: 151699]
- Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S. Autism Diagnostic Observation Schedule. Western Psychological Services; Los Angeles, CA: 2001.
- Lovaas OI, Koegel RL, Schreibman L. Stimulus overselectivity in autism: A review of research. *Psychological Bulletin*. 1979; 86:1236–1254. [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. [PubMed: 5556929]
- Mackie R, Mackay CK. Attention vs. retention in discrimination learning of low-MA retarded adults and MA-matched retarded children. *American Journal of Mental Deficiency*. 1982; 86:543–547. [PubMed: 7124807]
- Matthews B, Shute R, Rees R. An analysis of stimulus overselectivity in adults with autism. *Journal of Intellectual & Developmental Disability*. 2001; 26:161–176.
- Mayer-Johnson. Picture communication symbols: Boardmaker version 6 [Computer software]. Mayer-Johnson; Solana Beach, CA: 2008.
- Ploog B. 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. [PubMed: 20238154]
- Reed P. Brief Report: The effect of delayed matching to sample on stimulus over-selectivity. *Journal of Autism and Developmental Disorders*. 2012; 42:1515–1519. [PubMed: 21969076]
- Reed P, Petrina N, McHugh L. Over-selectivity as a learned response. *Research in Developmental Disabilities*. 2011; 32:201–206. [PubMed: 20951544]
- Reed SR, Stahmer AC, Suhrheinrich J, Schreibman L. Stimulus overselectivity in typical development: Implications for teaching children with autism. *Journal of Autism and Developmental Disorders*. 2013; 43:1249–1257. [PubMed: 23065100]
- Rieth SR, Stahmer AC, Suhrheinrich J, Schreibman L. Examination of the prevalence of stimulus overselectivity in children with ASD. *Journal of Applied Behavior Analysis*. 2015; 48:71–84. [PubMed: 25418837]
- Rincover A, Feldman M, Eason L. “Tunnel vision”: A possible keystone stimulus control deficit in autistic children. *Analysis and Intervention in Developmental Disabilities*. 1986; 6:283–304.
- Schneider HC, Salzberg CL. Stimulus overselectivity in a match-to-sample paradigm by severely retarded youth. *Analysis and Intervention in Developmental Disabilities*. 1982; 2:273–304.
- Schopler, E.; Van Bourgondien, ME.; Wellman, GJ.; Love, SR. Childhood Autism Rating Scale. 2nd ed.. Western Psychological Services; Los Angeles: 2010.
- Schreibman L, Charlop MH, Koegel RL. Teaching autistic children to use extra-stimulus prompts. *Journal of Experimental Child Psychology*. 1982; 33:475–491. [PubMed: 7097156]
- Selzter MM, Abbeduto L, Krauss MW, Greenburg J, Swe A. Comparison groups in autism family research: Down syndrome, fragile X syndrome, and schizophrenia. *Journal of Autism and Developmental Disorders*. 2004; 34:41–48. [PubMed: 15098956]
- Smith T, Eikeseth S. O. Ivar Lovaas: Pioneer of applied behavior analysis and intervention for children with autism. *Journal of Autism and Developmental Disorders*. 2011; 41:375–378. [PubMed: 21153872]
- Weigelt S, Koldewyn K, Kanwisher N. Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience and Biobehavioral Reviews*. 2012; 36:1060–1084. [PubMed: 22212588]
- Whiteley JH, Zaparniuk J, Asmundson GJG. Mentally retarded adolescents' breadth of attention and short-term memory processes during matching to sample discriminations. *American Journal of Mental Deficiency*. 1987; 92:207–212. [PubMed: 3434592]

Figure 1.

A, B, C: Baseline stimuli for Faces test. D: Example test stimulus for Faces test; Face B altered to have the eyes of Face A. E: Examples of stimuli for the Unfamiliar Forms test.

Figure 2.

Group mean accuracy scores for Color/form Compounds, Printed Word, Faces, 2 and 3 Color Symbols, and 2 and 3 Unfamiliar Forms Test phases; see Table 3 for standard deviations. Aut = Autism group, DS = Down syndrome group, TDC = typically developing children group.

Table 1

Numbers of Participants in Overselectivity Baseline and Test Phases

	Autism	DS	TDC
Total Participants	29	23	31
Color/form Compounds			
Start Baseline	12	13	12
Fail Baseline	0	0	0
Lost Stimulus Control	0	1	0
Completed Test	12	12	12
Printed Words			
Start Baseline	12	14	13
Fail Baseline	0	1	1
Lost Stimulus Control	0	1	0
Completed Test	12	12	12
Faces			
Start Baseline	21	23	23
Fail Baseline	7	8	10
Lost Stimulus Control	1	2	0
Completed Test	13	13	13
Color Symbols			
Start Baseline	14	14	14
Fail Baseline	0	0	0
Lost Stimulus Control	0	0	0
Completed Test	14	14	14
Unfamiliar Forms			
Start Baseline	26	22	24
Fail Baseline	0	0	0
Lost Stimulus Control	4	0	2
Completed Test	22	22	22

Table 2

Participant Demographic and Matched Variables Analysis for Overselectivity Tests

	Gender		CA (years)		DAS-II General Conceptual Ability Score		PPVT-4 Growth Scale Value Score		DAS-II Matrices Ability Score		
	N	M	F	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Color/form Compounds											
Autism	12	9	3	17.1	3.3	46.5	10.5	129.4	17.5	57.6	21.2
	12	8	4	15.6	2.9	41.0	8.1	134.0	21.6	50.8	12.9
	12	5	7	4.1	0.8	115.3	14.4	140.0	16.0	50.4	11.3
	Test		p = .21 ^a		p = .16 ^b		p = .24 ^b		p = .38 ^c		p = .47 ^c
Printed Words											
Autism	12	9	3	17.0	3.3	48.4	8.6	133.0	13.1	61.2	15.2
	12	8	4	14.9	2.3	41.4	7.2	140.0	18.9	52.0	11.1
	12	6	6	5.2	0.6	107.5	10.3	143.6	13.1	57.9	9.7
	Test		p = .43 ^a		p = .04 ^b		p = .06 ^b		p = .24 ^c		p = .19 ^c
Faces											
Autism	13	11	2	18.0	1.8	43.7	7.5	134.9	16.3	58.1	15.1
	13	7	6	14.5	2.8	42.9	8.1	143.4	16.6	52.0	12.2
	13	8	5	5.0	0.7	109.8	10.2	149.7	15.0	54.3	8.6
	Test		p = .22 ^a		p < .001 ^b		p = .80 ^b		p = .07 ^c		p = .44 ^c
Color Symbols											
Autism	14	12	2	17.9	3.2	47.1	10.2	130.6	18.5	55.9	17.3
	14	7	7	16.3	2.5	37.8	7.3	137.0	18.1	51.0	10.3
	14	7	7	5.1	0.4	108.9	13.7	141.8	9.6	55.1	10.0
	Test		p = .08 ^a		p = .07 ^b		p = .03 ^b		p = .19 ^c		p = .57 ^c
Unfamiliar Forms											
Autism	22	17	5	17.5	2.8	45.0	10.4	131.1	17.9	58.2	19.8
	22	12	10	15.2	3.1	39.4	7.9	136.0	21.0	48.7	11.3

Gender		CA (years)		DAS-II General Conceptual Ability Score		PPVT-4 Growth Scale Value Score		DAS-II Matrices Ability Score		
N	M	F	Mean	SD	Mean	SD	Mean	SD	Mean	SD
22	9	13	4.7	0.8	113.1	11.5	142.8	13.1	54.1	9.9
TDC		p = .05 ^a		p = .002 ^b		p = .07 ^b		p = .09 ^c		p = .10 ^c
Test										

Note. DS = Down syndrome. TDC = typically developing children. CA = chronological age in years. DAS-II = Differential Abilities Scales II; General Conceptual Ability Score is a standardized score, mean = 100, SD = 15. PPVT-4 = Peabody Picture Vocabulary Test 4. Tests for group differences:

^aGender = chi-square statistic.

^bCA and DAS-II GCA (Autism vs. DS only) = t-test.

^cPPVT-4 Growth Scale Value and DAS-II Matrices ability scores = ANOVA.

Table 3

Overselectivity Test Results: Accuracy Scores and Group Difference Testing

	Accuracy (% correct)		Group Test	Post-hoc Comparisons						
	Mean	SD		P	Autism vs. DS		Autism vs. TDC		DS vs. TDC	
					p-value ^b	ES	p-value ^b	ES	p-value ^b	ES
Color/form Compounds			.13 ^a	1.00 (.86)	.32	.23 (.07)	.75	.32 (.11)	.63	
Autism	97.7	1.0								
DS	97.4	0.6								
TDC	95.1	1.2								
Printed Words			.29	1.00 (.97)	.00	.48 (.16)	.70	.52 (.17)	.62	
Autism	91.7	9.0								
DS	91.5	11.4								
TDC	86.1	7.9								
Faces			.69 ^a	1.00 (.98)	.00	1.00 (.55)	.22	1.00 (.56)	.19	
Autism	85.9	8.3								
DS	85.8	10.9								
TDC	83.6	9.8								
Color Symbols 2 samples			.36	.78 (.26)	.45	1.00 (.85)	.00	.57 (.19)	.48	
Autism	82.6	10.7								
DS	78.0	11.3								
TDC	83.4	9.7								
Color Symbols 3 samples			.45	1.00 (.60)	.23	1.00 (.46)	.28	.63 (.21)	.56	
Autism	68.8	13.0								
DS	66.3	13.4								
TDC	71.9	7.8								
Unfamiliar Forms 2 samples			.44 ^a	.67 (.22)	.45	1.00 (.98)	.00	.64 (.21)	.42	
Autism	70.5	10.9								
DS	66.1	11.4								
TDC	70.6	12.9								
Unfamiliar Forms 3 samples			.56 ^a	.87 (.29)	.35	1.00 (.58)	.19	.33 (.11)	.45	
Autism	57.1	8.4								
DS	53.9	9.0								
TDC	58.8	12.6								

Note. DS = Down syndrome. TDC = typically developing children. ES = effect size (d).

^a Kruskal-Wallis used in lieu of ANOVA because violation of normality.

^b p values are Bonferroni adjusted for multiple comparisons; unadjusted p in parentheses.