J Autism Dev Disord (2013) 43:2048–2061
DOI 10.1007/s10803-012-1754-y

ORIGINAL PAPER

Residual Difficulties with Categorical Induction in Children with a History of Autism

Letitia R. Naigles · Elizabeth Kelley · Eva Troyb · Deborah Fein

Published online: 16 January 2013

Abstract In two experiments, typically developing (TD) children, high-functioning children with autism (HFA) and children with a history of autism who have achieved optimal outcomes (OOs), matched on age ($M = 13$ years) and nonverbal IQ, were asked to extend properties of categories to new items (categorical induction). All groups demonstrated some knowledge of category structure by extending at above-chance levels; however, the TD group extended more consistently than the OO and HFA groups. More consistent extenders had higher lexical and nonverbal IQ scores (Experiment 1) or higher pragmatics scores (Experiment 2). Thus, even very high functioning individuals with autism, or with an OO, still exhibit residual difficulties with category knowledge and extension; moreover, category tasks relate to a variety of verbal and nonverbal abilities. The difficulty these groups had with categorical induction may be related to their difficulty with generalization more widely; future research should investigate this possibility.

Keywords Optimal outcome · Categorical induction · Semantics · Pragmatics

L. R. Naigles (✉) · E. Kelley · E. Troyb · D. Fein
Department of Psychology, University of Connecticut, 406 Babbidge Road U-1020, Storrs, CT 06269-1020, USA
e-mail: letitia.naigles@uconn.edu

Present Address:
E. Kelley
Queen's University, Humphrey Hall, Room 351, 62 Arch St., Kingston, ON K7L 3N6, Canada

Introduction

For at least the past 20 years, researchers and clinicians have recognized that individuals with Autism Spectrum Disorders (ASDs) have impairments in semantic, lexical, and/or category organization; that is, in the way(s) that words and their meanings are connected, structured, and accessed in mental representations. These impairments can be observed across the lifespan, in preschoolers (Tek et al. 2008), in grade-schoolers and adolescents (Church et al. 2010; Dunn and Bates 2005; Kamio et al. 2007), and in adults (Dunn et al. 1996; Minshew et al. 2002). However, such findings are not universal—at least two recent papers have reported evidence for some *un*impaired lexical/semantic organization (Gastgeb et al. 2006; Norbury 2005). These discrepant findings may be attributed to at least two possibilities: first, to differences in the semantic specificity of the tasks, and second, to differences in the ASD samples under investigation, as the variability of this population, including its language abilities, is well known (Kelley 2011; Kjelgaard and Tager-Flusberg 2001; Tager-Flusberg and Joseph 2003). In the current studies, we investigate the first possibility by using a new (to this population) assessment of category organization, namely, categorical induction (Gelman 2003). This class of tasks taps how individuals determine whether a new exemplar carries the properties of its category or class, and the specific task used here has been shown to develop in typically developing (TD) children during the grade school years (Gutheil and Gelman 1997; Rhodes et al. 2008). We also investigate the second possibility, by comparing high functioning individuals with autism (HFA) and TD individuals, with a newly-described and recognized group, namely, optimal outcome children (OO). Our OO group is comprised of individuals who were diagnosed with an ASD before the age of 5 years, and who no longer—at the point of our assessments—show any significant autism symptoms (Fein et al. 2005, 2013; Helt et al. 2008; Kelley et al. 2006, 2010; Sutera et al. 2007). Including individuals with an OO allows us to investigate whether children's early difficulties with semantic organization can be overcome with development and/or intervention, as well as to investigate in more detail which aspects of functioning might be related to categorization strengths and weaknesses.

Lexical-Semantic Organization in Individuals with ASD

Dunn and her colleagues have examined semantic category structures by asking children with ASD to produce or respond to a series of animal and vehicle names; they found that grade-schoolers produced more atypical items (e.g., *yak, ocelot*) than TD peers, and did not show the usual 'deviance-detecting' brain response (i.e., N400 in event-related potentials (ERP) when hearing out-of-category names) (Dunn et al. 1996; Dunn and Bates 2005). These studies demonstrate that individuals with ASD may not have the same organization of categories as typically-developing individuals. That is, their semantic categories may include atypical and not prototypical members.

One of the most important factors in being able to categorize is being able to form the idea of what a prototypical member of the category is like; that is, most North Americans would assert that a robin is a prototypical bird, while a kiwi is not. Much research has shown that typically developing individuals implicitly extract prototypical features when learning a new category (Deng and Sloutsky 2012; Klinger and Dawson 2001). In contrast, Church et al. (2010) asked 9-year-old children with ASD to distinguish prototypical (i.e., including most of the relevant features or properties) versus outlier dot patterns (in the context of a game involving recognition of a 'cave ghost'), and found these children rejected the prototypical exemplars as members of the category more frequently than their TD peers. This study suggests that children with ASD have difficulties extracting similar features across exemplars, and/or using the frequency of occurrence of said features and exemplars, to form family resemblance or prototype semantic structures. More recently, Alderson-Day and McGonigle-Chambers (2011) also demonstrated that adolescents with ASD used category reasoning significantly less than TD controls in a Twenty Questions task; that is, they asked fewer abstract questions, and asked questions that eliminated fewer items (although this effect could be partially accounted for by group differences in verbal IQ). Furthermore, Kamio et al. (2007) have conducted semantic priming studies, in which individuals with ASD were shown pairs of letter strings and asked to judge whether the second string in each pair was a word or non-word; these targets were either: semantically related (*girl–boy*), phonologically related (*girl–curl*), or unrelated (*girl–chair*) to the first word (prime) in the pair. These adolescents showed no speeded responding for the semantically related targets over the others, although their TD peers did (see also Kamio and Toichi 2000). As with the more explicit tasks above, these latter findings indicate that children with ASD have difficulty forming semantic connections between lexical items, connections that would enable the construction of categories.

However, the findings of Norbury (2005) and Gastgeb et al. (2006) are somewhat at odds with this conclusion. Norbury asked school-age and adolescent children with ASD to match ambiguous words (e.g., *bank*) to images depicting either their dominant (financial institution) or subordinate (river's edge) meaning. She found low error rates—but also slower response times—for the subordinate matches for both the ASD and language-matched TD groups, indicating recognition of both meanings but also a *stronger* representation for the dominant meaning. Norbury next gave the same children the same ambiguous words, now embedded in neutral or biased contexts (e.g., the boy *ran* vs. *fished* from the bank), to match with the images, and found that the biased context increased accuracy and decreased response times for both groups. Such context effects indicate that the children recognized—and so had formed—semantic relationships between the biasing verbs and target nouns. Similar conclusions can be drawn from Gastgeb et al.'s (2006) study, in which they asked grade-schoolers and adolescents with ASD to judge whether images of typical, somewhat typical, and atypical exemplars were instances of a named category (e.g., *cat* or *chair*). All participants were less accurate with the atypical items (although responding correctly at above-chance levels: 92–94 % correct for atypical vs. 96–98 % correct for typical); moreover, the grade-schoolers—both TD and HFA—responded more slowly to the atypical than typical items (22 and 42 % more slowly, respectively), and the adolescents in both groups also responded more slowly to the somewhat typical than typical items. Finally, both HFA groups responded more slowly to the atypical items than the TD groups did. The slower and less accurate responses of the ASD groups to the atypical items suggests the presence of some category structure, as they were clearly distinguishing instances based on relative familiarity and/or similarity with some stored representation.

How can we make sense of these discrepant findings? One possibility is that individuals with ASD's performance with categorization is dependent on the level of complexity being tapped. For example, Gastgeb et al.'s (2006) task, which did find category sensitivity, may be considered to tap a 'simpler' type of category structure which has already been demonstrated in individuals with autism (e.g. Tager-Flusberg 1985). In contrast, the task of Kamio et al. (2007), which did not find category sensitivity, used a much wider range of items and tapped both close semantic (e.g., bird-nest) and far semantic (e.g., worm-nest) associates, thus requiring the participants to differentially activate more subtle aspects of semantics. However, Norbury's (2005) task might also be considered to tap complex and subtle semantic relations (e.g., distinguishing *bank/fish* and *bank/steal*), albeit here between verbs and nouns rather than among nouns, and still yielded comparable sensitivity by the ASD and TD groups. In sum, it is not yet clear whether a task- or level-of-semantics based explanation can successfully distinguish between those studies that demonstrate category sensitivity in children with ASD, and those that do not.

A second possible resolution of these discrepant findings might be traced to differences among the samples of children with ASD tested. For example, scrutiny of the standard scores reported for the studies of Dunn and colleagues reveals that the participating children had standard scores from language assessments in the 85–95 range (Dunn et al. 1996; Dunn and Bates 2005), whereas the participating children in Gastgeb et al. (2006) and Norbury (2005) had standard scores from language assessments in the 100–105 range. Nonverbal IQ scores seem to vary less across studies, clustering around 100 (although Alderson-Day and McGonigle-Chambers' (2011) participants also had VIQ scores around 105). Specific language scores are not reported for both TD and ASD groups in the Church et al. (2010) and Kamio et al. (2007) studies. It seems possible, then, that the semantics- and category-sensitive performance of the participants in the Gastgeb et al. (2006) and Norbury studies can be traced to their very high level of language functioning. Supporting this possibility, Norbury (2005) also included a group of children with ASD in her tasks whose language scores were much lower (i.e., around 80), but whose nonverbal IQ scores were comparable, and found much less evidence of contextual facilitation in these children. In sum, the ability of children with ASD to form complex categorical and semantic structures may be tied to their achieving a *high* level of general language; 'merely' testing above 80 (i.e., in the normal range) may not be sufficient.

The Current Studies: A New Task and a New Population

In the current set of studies, we explore this 'level of language' hypothesis in three ways. First, we consider the extent to which divergent findings on category formation and structure in ASD could be attributable to task variation, by introducing a task that assesses a different aspect of category structure. Categorical induction is the process by which people decide, without explicit instruction, whether specific properties observed for some exemplars of a category can also be extended to new items (Gelman 2003; Osherson et al. 1990). In general, if the properties are considered essential (rather than incidental) to the category, then they should be extended to new instances of that category. And while perceptual similarity between items is one marker of category membership, a more compelling marker is the sharing of a name. Gelman and her colleagues (e.g., Gelman and Markman 1986; see Gelman 2003 for a summary) have conducted extensive investigations into the developmental and conceptual characteristics of categorical induction; for example, they have taught TD toddlers and preschoolers a property for a given sample (object or animal; "here is a brown rabbit. It eats grass"), and then asked whether this property is extended to new targets with the same name (e.g., a white rabbit) versus similar surface characteristics (e.g., a brown squirrel). Gelman and colleagues have demonstrated that TD children as young as 2 years of age preferentially extend properties to targets given the same name as the sample over targets with similar surface characteristics. In the current studies, we investigate whether school-age and adolescent children with HFA (8–17 years), whose language and nonverbal IQ levels vary widely *within* the normal range (e.g., 80–130), are also able to extend properties to new members of a category.

Second, we include a new group of school-age and adolescent children, called optimal outcome (OO); these children were diagnosed with autism as toddlers or preschoolers and have now achieved an excellent outcome, such that their standardized language and adaptive behavior scores are well within the normal range (not differing significantly from TD agemates); also, diagnostic tests reveal that they no longer fall on the autism spectrum (Fein et al. 2005, 2013; Helt et al. 2008; Kelley et al. 2006, 2010; Sutera et al. 2007). Because this group has been sometimes found to perform at levels numerically below, albeit not significantly below, those of TD controls (Kelley et al. 2006, 2010), the OO group is likely to provide a 'middle ground' of variability between the TD and HFA groups. Indeed, in our previous research, we gave this task to a group of 7-year-old OO children, who extended preferentially by name numerically less frequently than their TD peers; this difference was statistically significant only for the animate items (Kelley et al. 2006). Thus, we have preliminary evidence that categorical induction tasks tap aspects of category structure that differentiate children with a history of autism from those who are typically developing.

Relatedly, our third tack for exploring the 'level of language' contribution to categorization abilities involves comparing the children's categorization performance with other measures of their language, including their vocabulary, grammar, and pragmatics. There have been several proposals recently that some individuals on the autism spectrum have no language delay and/or impairment whereas others have an impairment that is specific to certain areas of lexicon and grammar, similar to individuals with Specific Language Impairment (SLI) (Bishop and Norbury 2002; Kjelgaard and Tager-Flusberg 2001; Tager-Flusberg and Joseph 2003). To the extent that categorization performance correlates primarily with vocabulary or grammar measures in our studies, we may adduce support for these proposals; in contrast, categorization performance that correlates primarily with pragmatics and/or nonverbal measures may be an indication of a more pervasive difficulty in autistic functioning, possibly akin to weak central coherence (Happe and Frith 2006).

In the current studies, we assess a more 'advanced' aspect of category induction, namely, the preference for diverse sample sets over homogeneous or single sample sets in selecting which properties to extend to a target. That is, consider the situation in which a new animal exemplar is presented (say, a brown snake), and the task is to figure out whether it has blue eyes or black eyes. One array of snakes (a diverse array consisting of snakes of many different surface characteristics, none of them brown) all has blue eyes whereas another array of snakes (a homogeneous array of green snakes) all has gray eyes (see Fig. 1). Which property should be attached to the brown snake, given that it doesn't perfectly match any of those in either array? According to Osherson et al. (1990), the diverse array has more *coverage*; that is, it is more representative of the category as a whole. Therefore, the property belonging to the diverse array is a better choice for the target snake. Gelman and colleagues have performed this sort of task with TD children and adults (e.g., Gutheil and Gelman 1997; Rhodes et al. 2008), and found that adults, and children as young as 9 years of age, robustly prefer to extend the property of the diverse array over that of a homogeneous array or of a single item. In contrast, children under 8 years of age show a more fragile and easily mutable 'diversity preference.' Thus, we conjecture that children with weaker categorization abilities, such as those with ASD, may also have more difficulty demonstrating a diversity preference.

In Experiment 1, we investigated the category inductions of three groups of young adolescents (HFA, OO, and TD), asking them to extend the physical properties of arrays of familiar animals to a new target animal. Given previous findings indicating some category knowledge, we expected that these high-functioning individuals would show a preference for extending the properties of the diverse arrays; however, we also expected less consistent diversity preferences from the HFA—and possibly OO—children than from the TD children. These predictions being borne out, in Experiment 2 we studied somewhat younger children, and asked them to extend the *conceptual* properties of *unfamiliar* (i.e., novel) animate characters. If the diversity preference is stable, then it should be observed even with these more challenging stimuli. Within both experiments, we also compared the children's diversity preference scores to their concurrent language and non-verbal IQ scores, as well as their adaptive functioning scores, to ascertain which factors were most closely connected to their categorization abilities.

Experiment 1

Methods

Participants

Participants included 19 adolescents with a history of ASD who achieved optimal outcomes (OO), 26 high-functioning adolescents with a current ASD diagnosis (HFA), and 23 typically developing peers (TD). The participants in the study ranged from 9 years, 7 months to 17 years, 11 months. The three groups were matched on age, and gender. Groups did not differ on nonverbal IQ ($p > .50$) but were significantly different on verbal IQ, $p = .036$, with the HFA group scoring 7 points lower than the OO group and 10 points lower than the TD group. See Table 1 for participant characteristics. The participants were predominantly Caucasian, with only three individuals in the OO group, one individual in the HFA, and three individuals in the TD group reporting other races or ethnicities. Participants were recruited through flyers and information distributed to New England autism associations, advertisements posted in newspapers and online forums, and presentations at conferences. Participants were also referred from the principal investigators' private practice, the Psychological Services Clinic at the University of Connecticut, and from other ongoing studies. TD participants were additionally recruited through advertisements posted at local public schools and at the University of Connecticut. The study was approved by the Institutional Review Board.

Enrollment Criteria To be included in the study, all participants were required to have verbal, nonverbal, and full-scale IQ standard scores greater than 77 (within one-and-a-half standard deviations of the mean IQ of 100). The Wechsler Abbreviated Scale of Intelligence (WASI) was administered during testing in order to confirm IQ for all participants. Other eligibility requirements applied specifically to the separate participant groups, as described below.

Fig. 1 Example stimuli contrasting diverse and homogeneous arrays of a category

To be included in the OO group:

1. Participants had to have a documented history of an ASD diagnosis made by a specialist in the field of autism. Parents of participants were required to provide a written report that described an ASD diagnosis made before the age of 5. Early language delay (no words by 18 months or no phrases by 24 months) documented in the report was required. As a second step in confirming diagnosis, the report was edited to remove information about diagnosis, summary, and recommendations but leaving descriptions of behavior. One of the co-investigators (MB), an expert in diagnosis of ASD and Director of the

Table 1 Mean age, gender and standardized assessments of the groups, Experiment 1

	TD group N = 23	OO group N = 19	HFA group N = 26	*F* test, significance, effect size and group comparison
Age	12.83 years (1.57)	12.59 years (2.95)	12.77 years (2.41)	$F(2,64) < 1$
	Range = 9–15 years	Range = 9–17 years	Range = 9–17 years	Partial $\eta^2 = .002$
Gender	20 boys, 3 girls	14 boys, 5 girls	24 boys, 2 girls	$\chi^2(2) = 3.47$, ns
WASI	114.73 (12.39)	109.26 (13.00)	111.96 (14.7)	$F(2,64) < 1$
Performance standard score	Range = 89–139	Range 87–131	Range 78–147	Partial $\eta^2 = .02$
	N = 22		N = 25	
WASI	113.31 (13.43)	110.11 (15.52)	102.72 (13.48)	$F(2,64) = 3.52$
Verbal standard score	Range = 93–138	Range = 91–137	Range = 81–133	Partial $\eta^2 = .10$
	N = 22		N = 25	TD > HFA*
WASI	115.14 (11.84)	110.79 (13.93)	108.04 (12.99)	$F(2,64) = 1.78$
Full-scale standard score	Range = 97–142	Range = 82–134	Range = 80–138	Partial $\eta^2 = .054$
	N = 22		N = 25	
PPVT	121.23 (10.99)	112.32 (13.25)	108.88 (15.26)	$F(2,63) = 5.23$
Standard score	Range = 104–143	Range = 86–130	Range = 81–130	Partial $\eta^2 = .14$
	N = 22			TD > HFA*

University of Connecticut Psychological Services Clinic, reviewed these reports, blind to early diagnosis and current group membership. In addition to potential OO participants, she reviewed 24 "foil" reports for children with non-ASD diagnoses, such as global delay or language disorder. Four potential OO participants were rejected for insufficient early documentation, and were dropped from the study. All 24 foils were correctly rejected.

2. Participants could not currently meet criteria for any ASD according to the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2002) administered by a research-reliable interviewer. In addition, the ADOS of all potential OO cases was reviewed by a clinician with more than 15 years of autism diagnostic experience (IME, MB, or DF) who confirmed that ADOS scores were below ASD thresholds and that in their expert clinical judgment, an ASD was not present.
3. Participants had to perform in the average range or higher on a standardized measure of adaptive functioning. Specifically, participants' scores on the communication and socialization domains of the *Vineland Adaptive Behavior Scales* had to be greater than 77 (within one-and-a-half standard deviations of the mean IQ of 100).
4. Participants had to be fully included in regular education classrooms with no one-on-one assistance and no special education services to address autism deficits (e.g., no social skills training). However, participants in this group could be receiving limited special education services to address impairments not specific to ASDs, including language deficits, learning disorders, and psychiatric disorders.

To be included in the HFA group:

1. Following Collaborative Programs of Excellence in Autism diagnostic guidelines (Luyster et al. 2005), participants had to meet criteria for ASD on the ADOS (both Social and Communication domains and total score) and according to best estimate clinical judgment.

To be included in the TD group:

1. Participants could not meet criteria for any ASD at any point in their development, by parent report.
2. Participants could not have a first-degree relative with an ASD diagnosis
3. Participants could not meet current diagnostic criteria for an ASD on the ADOS or by clinical judgment. There was no attempt to exclude TD children for other learning or psychiatric disorders (but see general exclusion criteria).
4. As with the OO group, participants had to perform in the average range or higher on a standardized measure of adaptive functioning. Specifically, participants' scores on the communication and socialization domains of the *Vineland Adaptive Behavior Scales* had to be greater than 77 (within one-and-a-half standard deviations of the mean IQ of 100).

Exclusion Criteria Potential participants for any group were excluded from the study if (1) at the time of the telephone screening they exhibited symptoms of major psychopathology (e.g., active psychotic disorder) that would impede full participation, (2) they had severe visual or hearing impairments, or (3) they had a history of seizure disorder, Fragile X syndrome, or significant head trauma with loss of consciousness.

Procedure

The evaluation was administered over the course of two or three testing sessions in a quiet room with one examiner and lasted approximately 6 h. In most cases, parent interviews were conducted concurrently by a second examiner and lasted approximately 3 h for the OO and HFA groups and one-and-a-half hours for the TD group. At the end of each testing session, the participant received a monetary incentive for participation.

Materials

To assess autistic symptomatology, the *Autism Diagnostic Interview-Revised (ADI-R)* (Lord et al. 1995) and the *Autism Diagnostic Observation Schedule-Generic (ADOS-G)* (Lord et al. 2002) were administered. For the ADOS, either Module 3 or 4 was used depending on the age of the participant.

To assess adaptive functioning, we administered the *Vineland Adaptive Behavior Scales* (Sparrow et al. 1984), which is a frequently used parent interview which addresses three major aspects of the child's adaptive functioning: *Communication, Daily Living Skills,* and *Socialization*. Three parents in the typically developing group, one in the OO group, and four in the HFA group did not wish to complete this interview (their children still remained in the study).

To assess standardized verbal and nonverbal performance, children were tested on the Wechsler Abbreviated Scales of Intelligence (WASI; Wechsler 2003), yielding measures of verbal, non-verbal, and full-scale IQ, and on the *Peabody Picture Vocabulary Test-Third Edition* (PPVT; Dunn and Dunn 1997), which assessed vocabulary knowledge.

The categorical induction task (Gutheil and Gelman 1997) investigated whether participants used sample size and diversity of a variety of animal categories to extend categorical properties. The experimenter introduced the task with the following story: "I went to the zoo last week and I have some pictures of the animals I saw there. I know about some of the animals because the man who works at the zoo told me about them, but I don't know about a lot of the other animals. I'm going to show you these pictures and you can help me answer some questions about them. Now remember, the questions are still going to be about things that we can't see in the pictures, but might still be there."

Three diverse/single (DS) and three diverse/homogeneous (DH) sets of pictures were shown to participants (see Fig. 1). In both types of sets, participants were shown five pictures of an animal of the same species, all of whom looked different; in the DS set this quintet was paired with a picture of a single, different-looking animal of the same species, and in the DH set the diverse quintet was paired with a second quintet of animals of the same species that looked very similar to each other. The participant was then asked whether another yet-again-different looking animal of the same species exhibited the characteristic shared with the diverse set of animals, or alternatively exhibited with the characteristic shared with the single or homogeneous set of animals.

For example, in one of the DS sets, participants were shown a page containing a picture of a single snake paired with a page containing five pictures of snakes, all of whom looked different from each other (i.e., diverse). They were told that "This snake (point to the single snake) has blue eyes while these snakes (point to the diverse set) all have gray eyes." They were then shown a picture of a still different-looking snake, and asked if this new snake "had blue eyes like this snake (point) or had gray eyes like these snakes (point)." Participants were given one point for each choice of the property of the diverse set. In the DH trials, participants were shown a page containing, for example, five pictures of snakes that were the same size and color (e.g., green, homogeneous), paired with a page containing five pictures of snakes who all looked different from each other (i.e., diverse). They were told that "These snakes (pointing to the green ones) all have blue eyes while these snakes (pointing to the diverse ones) all have gray eyes." They were then shown a picture of a still different-looking snake, and asked if this new snake "had blue eyes like these snakes (point) or had gray eyes like these snakes (point)." Participants were given one point for each choice of the property of the diverse set. Raw scores were summed for each set of trials (DH, DS); scores ranged from 0 to 3 for each set.

Results and Discussion

The children's standardized test scores are displayed in Tables 1 and 2. Regarding adaptive functioning, the OO group had significantly higher communication and socialization scores, and marginally higher daily living scores, than the HFA group, and the TD group had significantly higher communication, daily living, and socialization scores than the HFA group. Moreover, the TD group had significantly higher PPVT scores than the HFA group. On none of these measures did the TD and OO groups differ significantly.

Table 3 presents the scores for each set of trials for each group of participants. All three groups extended the property of the diverse sets more frequently than expected by chance (of 1.5) when these were paired with the homogeneous sets (TD: $t(22) = 15.94$, $p < .01$; OO: $t(18) = 5.43$, $p < .05$; HFA: $t(25) = 5.10$, $p < .05$). The extension preferences were similar but somewhat weaker when the diverse sets

Table 2 Adaptive functioning, Experiment 1

Test	TD group N = 23	OO group N = 18	HFA group N = 26	*F* test, significance, effect size and group comparison
Vineland Adaptive Behavior Scales	95.5 (8.3)	95.47 (11.5)	86.13 (12.29)	$F(2,64) = 5.83$
Communication standard score	Range 85–110	Range 79–109	Range 51–108	Partial $\eta^2 = .154$
				TD > HFA*
				OO > HFA*
Vineland Adaptive Behavior Scales	87.59 (7.94)	87.05 (15.78)	77.42 (14.01)	$F(2,64) = 4.44$
Daily living skills standard score	Range 74–102	Range 70–117$^+$	Range 46–110	Partial $\eta^2 = .122$
				TD > HFA**
				OO > HFA$^+$
Vineland Adaptive Behavior Scales	102.36 (8.06)	100.84 (8.37)	77.58 (15.16)	$F(2,64) = 35.39$
Socialization standard score	Range 91–117	Range 80–113	Range 47–109	Partial $\eta^2 = .525$
				OO > HFA***
				TD > HFA***

$^+$ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

were paired with the single pictures: for the TD group, $t(22) = 3.24$, $p < .05$, for the OO group, $t(18) = 1.85$, $p < .10$, and for the HFA group, $t(25) = 2.36$, $p < .05$. Moreover, the TD participants extended the property of the diverse sets over the homogeneous sets significantly more than both the OO and HFA participants [$t(40) = 2.72$, $p < .01$ and $t(47) = 2.49$, $p = .016$, respectively], who did not differ from each other. An ANCOVA was performed, co-varying nonverbal IQ, and a significant effect of group remained ($p = .048$). Because these scores were ordinal rather than continuous, we redid all of the statistical tests using arcsin-transformed scores, and obtained the same pattern of effects and non-effects.

Given that the general pattern of all three groups was to extend the property of the diverse set of items, we next investigated whether the groups differed in how consistently they performed this extension. Therefore, the participants were categorized as *perfect extenders* (correct extensions of the property of the diverse set for 6/6 trials), *almost perfect extenders* (5/6 trials), *consistent extenders* (4/6 trials), *moderate extenders* (3/6 trials) and *nonextenders* (0, 1, or 2 trials). The results for percentage of each group are shown in Table 4. A χ^2 test yielded a significant effect of the distribution of the percentage of participants in each category by group ($\chi^2(8) = 40.44$, $p < .001$); follow-up χ^2 tests revealed that the distribution of participants in the TD group was significantly different from that of the OO group ($\chi^2(3) = 17.295$, $p < .001$) and the HFA group ($\chi^2(4) = 27.65$, $p < .001$), which differed marginally from each other ($\chi^2(4) = 8.24$, $p = .083$).

Pairwise correlations were performed among all measures for all three groups combined. Significant relationships were obtained between the DH measure and the nonverbal IQ score ($r = .31$, $p = .012$), and the WASI full-scale IQ score ($r = .33$, $p = .007$); a marginally significant relationship was observed between the DH and verbal IQ measures ($r = .24$, $p = .056$). A significant relationship was also observed between the DHDS (combining the DH and DS scores) measure and the PPVT score ($r = .25$, $p = .038$).

In sum, when extending physical properties of animals to new instances, both HFA children currently on the ASD spectrum, and children who formerly carried an ASD diagnosis but currently manifested an optimal outcome, were able to recognize that the properties of the diverse arrays of animals were more reliable extensions than

Table 3 Mean categorical induction scores (SD), Experiment 1

Trials	Group		
	TD (n = 23)	OO (n = 18)	HFA (n = 28)
Homogeneous versus diverse[a,b]	2.83 (0.39)[c]**	2.37 (0.70)*	2.35 (0.85)*
Single versus diverse	2.00 (0.7)*	1.95 (1.03)$^+$	1.88 (0.82)*

[a] Range = 0–3

[b] Underlined choice is correct

[c] Differences from chance performance: $^+$ $p < .10$; * $p < .05$; ** $p < .01$

Table 4 Types of extenders by group (% of group), Experiment 1

Type of extender	Group		
	TD	OO	HFA
Perfect (6/6 trials)	17.3	16.7	15.4
Almost perfect (5/6 trials)	52.2	27.8	23.1
Consistent (4/6 trials)	26.1	38.9	38.4
Moderate (3/6 trials)	4.3	16.7	15.4
Nonextender	0	0	7.7

properties of homogeneous or single arrays of those animals. Hence, these children, as well as the TD group, performed above chance. However, both the HFA and OO groups nonetheless extended the properties of the diverse arrays significantly less consistently than the TD group, highlighting some residual category difficulties even with these very high-functioning individuals. Extending properties to diverse arrays was related to both verbal and nonverbal abilities in the children. Given the groups' overall good performance, we next asked whether such children could also preferentially target diverse over homogeneous or single arrays under more challenging conditions; that is, when the items were novel rather than familiar animals and the properties were conceptual (i.e., activity or food preferences) rather than physical. Rhodes et al. (2008) have reported that 6–8-year-old TD children's ability to rely on diverse arrays for induction diminishes under these conditions; we tested children with a mean age of 10 years (range 8–14 years; one 8-year-old in each group).

Experiment 2

Method

Participants

Three groups of children were examined in this study (see Table 5). After several children were eliminated for the reasons given below, the final sample included nine children with a history of ASD who had reached an optimal outcome level (the OO group), 11 typically developing children (the TD group), and 12 children on the autism spectrum who were of average intelligence (the HFA group).

The OO group was chosen a priori by a somewhat different set of criteria than in Experiment 1: (1) They had been mainstreamed into a regular classroom; (2) They had a full-scale IQ on their last school assessment which was greater than 70; (3) They were receiving no more than 1 h per week of service overall (e.g. 1 h of speech therapy, or 1 h of occupational therapy) and did not have an educational aide in the classroom; (4) They were considered by the school system to no longer be on the autism spectrum; (5) They had previously been diagnosed on the autism spectrum by a clinician who specialized in ASD and they met criteria on the *ADI-R* based on their 'ever' scores; (6) They no longer met criteria for an ASD diagnosis on the *ADOS-G* or for autism on the *ADI-R* current algorithm.

The HFA group was chosen a priori by the following criteria: (1) if they had been mainstreamed into a regular classroom, they continued to receive extra help. Two of the HFA children were being home-schooled, two were half-time in special needs classrooms and half-time mainstreamed with a one-on-one aide, three were mainstreamed with a one-on-one aide, and seven were mainstreamed with a shared aide. (2) They had a full-scale IQ on their last school assessment which was greater than 70; (3) They retained their classification of an ASD in the school system; (4) They had previously been diagnosed on the autism spectrum by a clinician who specialized in ASD and they met criteria on the *ADI-R* based on their 'ever' scores; (5) They currently met criteria for an ASD or autism diagnosis on the *ADOS-G* and the *ADI-R* current algorithm.

The typically developing children had no history of academic, neurological, or psychological problems as reported by their parents. Children in all of the groups were from middle- to upper-middle class families that resided in suburban and rural areas of the Northeastern US.

Seven children in the OO group in Experiment 2 had also participated in a previous study (Kelley et al. 2006), as did six children in the TD group. The testing sessions for Experiment 2 took place approximately 3 years after they had participated in this prior study (Kelley et al. 2006). Seven of these children also participated in Experiment 1 when they were older (one in the TD group, six in the HFA group). Information on the age, gender, and non-verbal IQ scores of the final sample of children can be found in Table 5. There were no significant differences between the groups on gender, age, or non-verbal intelligence as measured by the Matrix Reasoning subtest of the *Wechsler Intelligence Scale for Children-IV* (Wechsler 2003), although the HFA group did differ from the TD and OO groups on verbal IQ (see below and Table 6).

Materials

The *ADI-R, ADOS-G,* and *Vineland Adaptive Behavior Scales* were administered.

To assess general verbal functioning, the *Similarities, Vocabulary, and Comprehension* subscales of the *Wechsler Intelligence Scale for Children-Fourth Edition* (WISC; Wechsler 2003) were used and the Verbal Composite Index (equivalent to verbal IQ) was calculated. Children were

Table 5 Age, gender, non-verbal intelligence, and adaptive functioning, Experiment 2

	TD group N = 11	OO group N = 9	HFA group N = 12	*F* test, effect size, group comparison
Age	Mean 10 years, 2 months (17.5)	Mean 9 years, 11 months (15.1)	Mean 11 years, 4 months (26.1)	$F(2,31) = 1.901$
				ns
	Range 8–12 years	Range 8–12 years	Range 8–14 years	
Gender	8 boys, 3 girls	6 boys, 3 girls	7 boys, 5 girls	$\chi^2(2) = 0.864$, ns
Matrix	Mean = 12.27 (1.42)	Mean = 12.13 (2.75)	Mean = 9.82 (3.13)	$F(2,29) = 3.161^{+}$
Reasoning scaled score (NVIQ)	Range 12–16	Range 9–17	Range 4–14	Partial $\eta^2 = .19$
		N = 8		TD > HFA^{+}
VABS communication standard score	102.10 (8.72)	105.11 (13.99)	88.33 (11.09)	$F(2,30) = 6.765$ *
	Range 88–113	Range 86–122	Range 73–110	Partial $\eta^2 = .33$
				TD > HFA*
				OO > HFA*
VABS	98.20 (13.99)	93.89 (23.71)	77.83 (21.31)	$F(2,30) = 3.181^{+}$
Daily living skills standard score	Range 73–116	Range 55–132	Range 47–107	Partial $\eta^2 = .185$
				TD > HFA^{+}
VABS	107.10 (11.13)	98.00 (21.34)	69.00 (12.84)	$F(2,30) = 18.72$ **
Socialization standard score	Range 83–117	Range 53–127	Range 53–98	Partial $\eta^2 = .572$
				OO > HFA**
				TD > HFA**

$^{+}$ $p < .10$; * $p < .01$; ** $p < .001$

Table 6 Language and communication composite measures, Experiment 2

	TD group N = 11	OO group N = 9	HFA group N = 12	*F* test, significance, effect size and group comparison
PPVT	118.60 (13.31)	114.44 (17.401)	96.82 (16.27)	$F(2,29) = 5.713$*
Standard score	N = 10	Range 97–140	N = 11	Partial $\eta^2 = .297$
	Range 99–140		Range = 75–131	OO > HFA*
				TD > HFA**
TOPL	110.82 (8.43)	101.67 (8.42)	81.55 (16.41)	$F(2,30) = 15.465$**
Standard score	Range 91–126	Range 88–112	N = 11	Partial $\eta^2 = .525$
			Range 58–105	OO > HFA**
				TD > HFA**
WISC	124.27 (17.018)	111.11 (18.038)	90.64 (15.022)	$F(2,30) = 11.387$**
Verbal composite Index	Range 100–152	Range 79–140	N = 11	Partial $\eta^2 = .449$
			Range 55–119	OO > HFA*
				TD > HFA**
CELF	123.91 (18.163)	114.67 (22.006)	94.08 (18.023)	$F(2,31) = 7.238$**
Receptive language standard score	Range 92–150	Range 84–150	Range 53–120	Partial $\eta^2 = .333$
				TD > HFA**
TLC	113.33 (18.901)	98.25 (19.72)	80.08 (10.335)	$F(2,28) = 11.114$**
Interpreting intentions standard score	N = 9	N = 8	Range 65–94	Partial $\eta^2 = .461$
	Range 85–135	Range 65–118		OO > HFA^{+}
				TD > HFA**

$^{+}$ $p < .10$; * $p < .01$; ** $p < .001$

also tested on the *PPVT* (Dunn and Dunn 1997) to assess vocabulary knowledge, and the *Clinical Evaluation of Language Fundamentals Fourth Edition* (CELF; Semel et al. 2003) to assess both semantic and syntactic aspects of language. To assess pragmatic language ability, the *Test of Pragmatic Language* (TOPL: Phelps-Terasaki and

Phelps-Gunn 1992) was administered. All tests were administered with the standard procedure outlined in the manuals.

The categorical induction task in Experiment 2 investigated whether children used sample size and diversity of unfamiliar (nonsense) sentient character categories to extend their preferences for activities or food. For example, children were shown a number of nonsense characters called 'zugs' and told that these 'zugs' love peanut butter. They were then shown another 'zug' and asked if it loved peanut butter. Items varied according to whether the sample comprised one ($n = 4$) versus five ($n = 6$) characters; moreover, the multi-character items either looked the same (*homogeneous* e.g., all purple, $n = 3$) or differed in color (*diverse*; $n = 3$). This task further differed from that of Gutheil and Gelman (1997), and from Experiment 1, in that children were asked yes/no questions (Does this zug love peanut butter?) instead of forced-choice questions (Does this zug love peanut butter, like these zugs, or melon, like this zug?). This adaptation was made because the yes/no questions were shorter than the forced-choice question, and we thought the shorter questions would be easier for the children, as Gutheil and Gelman had found that TD eight-year-olds were just beginning to use the sample-diversity strategy. To guard against the possibility that the children in this study would get extremely good or poor scores via a yes-bias or no-bias, both 'liking' and 'not liking' were included. That is, six of the items involved things (e.g., food) or activities (e.g., riding bikes) the creature(s) did like, prompting an affirmative answer for categorical induction (i.e., if the zugs like peanut butter, then this new zug should also like peanut butter), and four involved things/activities the creature(s) did not like, prompting a negative answer (e.g., if the zugs do not like peanut butter, then this new zug should also not like peanut butter). Two 'single' items, one 'homogeneous' item, and one 'diverse' item involved negation.

The children were given a score of 1 for each answer in which they agreed that the new nonsense creature liked (or disliked) the thing/activity shared by other character(s) with the same label. Otherwise, they were given a score of zero. The proportion of correct answers was then compared across the trials in which the samples were homogeneous ($n = 3$) versus those in which the samples were diverse ($n = 3$), with a total possible score of 6.

Procedure

The children were administered the tasks during one of two sessions, as well as some additional experimental tasks that are not reported here (Kelley et al. 2010). It should be noted that where missing data occurred it was never because the child scored at floor; rather, in most instances the testing sessions ran too long and children could not complete the final test, or there was experimenter error or videotape malfunction.

Results and Discussion

The children's standardized test scores are displayed in Tables 5 and 6. These scores were discussed at length in Kelley et al. (2010); here, we note that this subset of the larger sample demonstrated the same group differences as reported earlier: Regarding adaptive functioning, the TD group had significantly higher communication scores than the HFA group, and both the TD and OO groups had significantly higher socialization scores than the HFA group. Regarding general language, both the TD and OO groups had significantly higher PPVT (lexical), TOPL (pragmatic), and WISC (verbal IQ) scores than the HFA group. The TD group had significantly higher scores on both CELF (semantics and structural language) measures than the HFA group. On none of these measures did the TD and OO groups differ significantly.

Figure 2 presents the mean percent correct extensions produced by each group for the diverse and homogeneous items. All three groups produced more correct extensions for the diverse arrays ($M = 78.13$ %, $SD = 37.49$) than for the homogeneous arrays ($M = 71.35$ %, $SD = 36.22$); this effect was significant with all groups combined ($F(31) = 4.122$, $p = .05$, partial $\eta^2 = .117$). Although the TD children produced across all items a numerically higher percentage of correct extensions ($M = 88.64$ %, $SD = 15.83$) than either the OO children ($M = 68.21$ %, $SD = 37.12$) or the HFA children ($M = 65.74$ %, $SD = 42.22$), differences among the three groups did not reach significance. However, only the TD group produced correct extensions more frequently than expected by a chance level of 50 % [diverse items: $t(9) = 8.79$, $p < .01$; homogeneous items: $t(9) = 4.72$, $p < .05$].

As in Experiment 1, the participants were categorized as *perfect extenders* (correct extensions in the DH and DS sets for 6/6 trials), *almost perfect extenders* (5/6 trials), *consistent extenders* (4/6 trials), *moderate extenders* (3/6 trials) and *nonextenders* (0, 1, or 2 trials). The results for percentage of each group are shown in Table 7. Most of the children in each group were extenders; however, 33 % of HFA children and 22.2 % of OO children extended very little. A χ^2 test yielded a significant effect of the distribution of the percentage of participants in each category by group ($\chi^2(8) = 66.48$, $p < .001$); follow-up χ^2 tests revealed that the distribution of participants in the TD group was significantly different from that of the OO group ($\chi^2(4) = 35.31$, $p < .001$) and the HFA group ($\chi^2(4) = 57.17$, $p < .001$), which differed significantly from each other ($\chi^2(3) = 13.57$, $p = .003$).

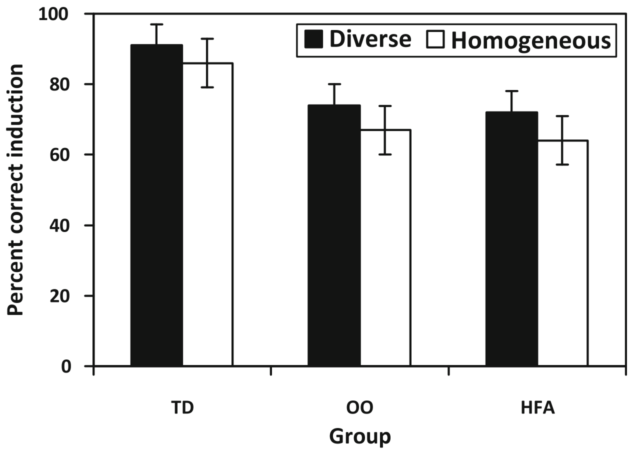

Fig. 2 Percent correct inductions produced by children (MA = 10 years) in Experiment 2, by group and type of array

Pairwise correlations were performed among all measures with all three groups combined. Variables that correlated significantly with the categorical induction measures are listed in Table 8. Because so many significant correlations emerged, partial correlations were then performed, controlling first for the children's Matrix Reasoning (nonverbal IQ) score, then for their PPVT score, and then for their TOPL score. When Matrix Reasoning was controlled, the other four variables (PPVT, WISC-VIQ, TOPL, CELF) still correlated significantly with the categorical induction measures. When PPVT was controlled, only the TOPL score still correlated significantly with the categorical induction measures, and when the TOPL score was controlled, no other variables remained correlated with the categorical induction measures. Therefore, it seems likely that the children's ability to extend properties—in this study, the preferences of novel animated characters—was largely a function of their knowledge of pragmatics.

General Discussion

These experiments investigated the degree to which children with a history of autism would preferentially extend the properties of *diversely* instantiated categories to new instances. In the first demonstration of this ability, good extension was observed for both children with HFA and those with an OO. Thus, these children exhibit well-formed category structure and sensitivity to both specifics of items and similarities across items. If the children were only sensitive to item characteristics, we would have seen more extensions from the homogeneous and/or single arrays than the diverse ones (i.e., they would have noticed the many different characteristics of the diverse arrays and decided that extension from these was unreliable). If the children were unwilling to extend at all, we would have seen chance responding in Experiment 1 (when they had to choose one of the arrays to extend by) and below-chance responding in Experiment 2 (when they had the option to refuse to extend at all). We have shown that children with a history of autism can preferentially extend the properties of diverse over single or homogeneous arrays when these arrays include novel as well as familiar items, and when they include conceptual as well as physical properties.

Table 7 Types of extenders by group (% of group), Experiment 2

Type of extender	Group		
	TD	OO	HFA
Perfect (6/6 trials)	63.6	44.4	41.6
Almost perfect (5/6 trials)	18.2	22.0	25.0
Consistent (4/6 trials)	9.1	0	0
Moderate (3/6 trials)	9.1	11.0	0
Nonextender	0	22.0	33.3

Nonetheless, we also found subtle deficits in both experiments, in both the HFA *and* OO groups; these children were not quite as consistent at extending as TD children (see Tables 4, 7). These findings corroborate Gastgeb et al. (2006; in press) and Alderson-Day and McGonigle-Chambers (2011), that individuals with ASD can acquire and use category structure, but do this less consistently than TD individuals. That is, the children and adolescents with ASD in these experiments were significantly less likely to extend the property of the diverse array in Experiment 1, and to extend properties at all in Experiment 2, than the TD participants. As predicted, the OO group performed 'in the middle', including more consistent extenders than the HFA group, but fewer than the TD group.

The ability to extend properties overall—both physical and preferences—clearly had a nonverbal reasoning component, as nonverbal ability emerged as a significant correlate in both experiments. However, language-related measures were also correlated with the children's extension performance. In Experiment 1, the ability to preferentially extend the physical properties of diverse arrays seemed linked to lexical aspects of language. In Experiment 2, the more challenging task of determining whether to extend the activity or food preferences of an unfamiliar category was correlated with all areas of language but seemed most closely tied to pragmatics because when the pragmatics measures were controlled, no other significant correlates remained. Possibly, because this experiment involved novel sentient characters, the ability to determine whether or not to extend their preferences also tapped the children's understanding of the social world, which is fragile even in HFA and OO groups.

To return to the questions raised in the introduction, the current findings suggest that both differences among ASD

Table 8 Significant pairwise correlations among variables, Experiment 2

Categorical induction measure	PPVT	Matrix	WISC-VIQ	CELF	TOPL
Percent diverse	.466**	.409*	.429*	.567**	.663***
Percent homogeneous	.634***	.409*	.497**	.567**	.615***

* $p < .05$; ** $p < .01$; *** $p < .001$

samples and task complexity are involved in semantic category formation and usage in individuals with ASD. Our HFA and OO groups, with language scores similar to those of the successful users of context and categorizers in Norbury (2005) and Gastgeb et al. (2006), performed above chance in Experiment 1, and included numerous consistent extenders in Experiment 2. In both experiments, those with higher general language/verbal IQ scores performed better than those with lower scores (albeit still within the normal range). Thus, better verbal ability evidently enabled children and adolescents with ASD to demonstrate this complex categorization ability. However, even some adolescents, with a history of ASD but currently showing an optimal outcome indistinguishable from TD controls on standardized tests, were still inconsistent in their category extensions. The complexity of the semantic task also seemed to play a role: Being asked to observe both the similarities and differences of the items in the arrays, perceive category structures, and then preferentially extend the nonvisible properties of the *diverse* sets to new instances seems especially challenging. Moreover, when the task involved novel sentient characters (Experiment 2), both lexical and pragmatic knowledge were evidently required. Thus, the ability to extend properties based on category knowledge requires more than just an excellent vocabulary.

Limitations and Future Directions

Limitations of these studies include the small sample size, especially in Experiment 2; thus, it would be beneficial to replicate both studies with a larger set of participants. Moreover, including a wider age range would enable an investigation of the degree to which the different relationships with categorical induction that were observed (nonverbal cognition and lexical in Experiment 1, nonverbal cognition and pragmatics in Experiment 2) are attributable to developmental factors (participants in Experiment 1 were older) and/or stimulus-related factors. Further research might also explore the connection between categorical induction performance and other difficulties observed in children with ASD, such as with narratives and social skills; to the extent that categorical induction is related to these types of pragmatics abilities, this type of generalization task may provide a clinical indicator, as well as an avenue for intervention. Another future direction, which we are currently pursuing, involves investigating precursors to categorical induction understanding; that is, what aspects of early development (e.g., nonverbal cognition, word learning biases, speed or organization of vocabulary growth) might be predictors of later facility with category organization and categorical induction?

In sum, category structure, particularly the ability to induce information from categories, is not completely missing in children with a history of autism, but neither is it completely intact. Tasks that target very specific aspects of category structure reveal surprising capacities, but also subtle deficits, in both children with HFA and with an OO. Moreover, as we surmised in the introduction, this ability is linked to children's level of language, both lexical and pragmatic. Categorical induction enables individuals to understand and make predictions about new items or events they encounter, and children with HFA manifest some inconsistencies with this process which may illuminate their attendant difficulties with novelty. Making accurate predictions of consequences for behavior is essential to navigating novel social interactions and inhibiting behaviors that may seem odd or inappropriate to peers. Our results also suggest that despite the remarkably good outcome of the OO group, they are not fully up to the TD controls in the cognitive processes tapped by these induction tasks. Since there are no standard scores for the induction tasks, however, it is quite possible that the OO or HFA groups' performance would be within the normal range, or even fully average (both groups performed above chance). The control group had above-average verbal and nonverbal IQ scores, making it more likely to find comparative deficits in the clinical groups. Since categorical induction is a fundamental cognitive process that helps with conceptual learning, standardizing induction and similar reasoning tasks would enable them to have clinical utility. That is, categorical induction tasks have the potential to illuminate how challenges in the semantic/reasoning domains demonstrated by children with ASDs might be connected to their challenges in the social domains. Developing a clinical measure of generalization useful at the individual level might add very significantly to our assessment and intervention planning armamentarium, since the ability to generalize appropriately underlies the effectiveness of most therapeutic teaching.

Acknowledgments This research was funded by R01 MH076189 from NIMH to D.F. and by R01 2DC007428 to L.R.N. We gratefully acknowledge the assistance of Molly Helt, Alyssa Orinstein, Michael

Rosenthal, and Katherine Tyson, and a number of UConn undergraduates, in conducting Experiment 1, as well as the children and families who participated. Preliminary results from these experiments were presented at the International Meetings for Autism Research in 2006 and 2009.

References

Alderson-Day, B., & McGonigle-Chambers, M. (2011). Is it a bird? Is it a plane? Category use in problem-solving in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 41*, 555–565.

Bishop, D. V. M., & Norbury, C. F. (2002). Exploring the borderlands of autistic disorder and specific language impairment: A study using standardised diagnostic instruments. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*, 917–931.

Church, B., Krauss, M., Lopata, C., Toomey, J., Thomeer, M., Coutinho, M., et al. (2010). Atypical categorization in children with high-functioning autism spectrum disorder. *Psychonomic Bulletin & Review, 17*, 862–868.

Deng, W. S., & Sloutsky, V. M. (2012). Carrot eaters or moving heads: Inductive inference is better supported by salient features than by category labels. *Pscyhological Science, 23*, 178–186.

Dunn, M., & Bates, J. (2005). Developmental change in neural processing of words by children with autism. *Journal of Autism and Developmental Disorders, 35*, 361–376.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test-third edition*. Circle Pines, MN: American Guidance Services.

Dunn, M., Gomes, H., & Sebastian, M. (1996). Prototypicality of responses in autistic, language-disordered, and normal children in a word fluency task. *Child Neuropsychology, 2*, 99–108.

Fein, D., Barton, M., Eigsti, I. M., Kelley, E., Naigles, L., Schultz, R., et al. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry, 54*, 195–205. doi:10.1111/jcpp.12037.

Fein, D., Dixon, P., Paul, J., & Levin, H. (2005). Brief report: Pervasive developmental disorder can resolve into ADHD: Case illustrations. *Journal of Autism and Developmental Disorders, 35*, 525–534.

Gastgeb, H., Dundas, E., Minshew, N., & Strauss, M. (in press). Category formation in autism: Can individuals with autism form categories and prototypes of dot patterns? *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-011-1411-x.

Gastgeb, H., Strauss, M., & Minshew, N. (2006). Do individuals with autism process categories differently? The effect of typicality and development. *Child Development, 77*, 1717–1729.

Gelman, S. A. (2003). *The essential child*. New York: Oxford University Press.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

Gutheil, G., & Gelman, S. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology, 64*, 159–174.

Happe, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*, 5–25. doi:10.1007/s10803-005-0039-0.

Helt, M., Kelley, E., Kinsbourne, M., Pandey, J., Boorstein, H., Herbert, M., et al. (2008). Can children with autism recover? If so, how? *Neuropsychology Reviews, 18*, 339–366.

Kamio, Y., Robins, D., Kelley, E., Swainson, B., & Fein, D. (2007). Atypical lexical/semantic processing in high-functioning autism spectrum disorders without early language delay. *Journal of Autism and Developmental Disorders, 37*, 1116–1122.

Kamio, Y., & Toichi, M. (2000). Dual access to semantics in autism: Is pictorial access superior to verbal access? *Journal of Child Psychology and Psychiatry, 41*, 859–867.

Kelley, E. (2011). Language in ASD. In D. Fein (Ed.), *The neuropsychology of autism* (pp. 123–137). New York, NY: Oxford University Press.

Kelley, E., Naigles, L. R., & Fein, D. (2010). An in-depth examination of optimal outcome children with a history of autism spectrum disorders. *Research in Autism Spectrum Disorders, 4*, 526–538.

Kelley, E., Paul, J., Fein, D., & Naigles, L. (2006). Residual language deficits in optimal outcome children with a history of autism. *Journal of Autism and Developmental Disorders, 36*, 807–828.

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes, 16*, 287–308.

Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology, 13*, 111–124.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2002). *Autism diagnostic observation schedule-generic*. Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., & LeCouteur, A. (1995). *Autism diagnostic interview-revised* (3rd ed., *Short Form*). Chicago: University of Chicago.

Luyster, R., Richler, J., Risi, S., Hsu, W., Dawson, G., Bernier, R., et al. (2005). Early regression in social communication in autism spectrum disorders: A CPEA study. *Developmental Neuropsychology, 27*, 311–336. doi:10.1207/s15326942dn2703_2.

Minshew, N., Meyer, J., & Goldstein, G. (2002). Abstract reasoning in autism: A dissociation between concept formation and concept identification. *Neuropsychology, 16*, 327–334.

Norbury, C. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autism spectrum disorders. *Journal of Experimental Child Psychology, 90*, 142–171.

Osherson, D., Smith, E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185–200.

Phelps-Terasaki, D., & Phelps-Gunn, T. (1992). *Test of Pragmatic Language*. Austin, TX: Pro-Ed.

Rhodes, M., Brickman, D., & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition, 108*, 541–556.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals* (4th ed.). San Antonio, TX: The Psychological Corporation.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Service, Inc.

Sutera, S., Pandey, J., Esser, E. L., Rosenthal, M. A., Wilson, L. B., Barton, M., et al. (2007). Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*, 98–107.

Tager-Flusberg, H. (1985). Basic level and superordinate level categorization by autistic, mentally retarded, and normal children. *Journal of Experimental Child Psychology, 40*, 450–469.

Tager-Flusberg, H., & Joseph, R. M. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society of London, 358*, 303–314.

Tek, S., Jaffery, G., Fein, D., & Naigles, L. R. (2008). Do children with autism show a shape bias in word learning? *Autism Research, 1*, 202–215.

Wechsler, D. (2003). *Wechsler intelligence scale for children-fourth edition*. San Antonio, TX: The Psychological Corporation.