JSLHR

Research Article

Gaze Toward Naturalistic Social Scenes by Individuals With Intellectual and Developmental Disabilities: Implications for Augmentative and Alternative Communication Designs

Jiali Liang[a] and Krista Wilkinson[a,b]

Purpose: A striking characteristic of the social communication deficits in individuals with autism is atypical patterns of eye contact during social interactions. We used eye-tracking technology to evaluate how the number of human figures depicted and the presence of sharing activity between the human figures in still photographs influenced visual attention by individuals with autism, typical development, or Down syndrome. We sought to examine visual attention to the contents of visual scene displays, a growing form of augmentative and alternative communication support.
Method: Eye-tracking technology recorded point-of-gaze while participants viewed 32 photographs in which either 2 or 3 human figures were depicted. Sharing activities between these human figures are either present or absent. The sampling rate was 60 Hz; that is, the technology gathered 60 samples of gaze behavior per second, per participant. Gaze behaviors, including latency to fixate and time spent fixating, were quantified.
Results: The overall gaze behaviors were quite similar across groups, regardless of the social content depicted. However, individuals with autism were significantly slower than the other groups in latency to first view the human figures, especially when there were 3 people depicted in the photographs (as compared with 2 people). When participants' own viewing pace was considered, individuals with autism resembled those with Down syndrome.
Conclusion: The current study supports the inclusion of social content with various numbers of human figures and sharing activities between human figures into visual scene displays, regardless of the population served. Study design and reporting practices in eye-tracking literature as it relates to autism and Down syndrome are discussed.
Supplemental Material: https://doi.org/10.23641/asha.6066545

Augmentative and alternative communication (AAC) often relies on visual supports for the development of receptive and expressive communication. When the visual supports involve tools or equipment such as pictures, line drawings, communication boards, or speech-generating devices, they are called aided AAC. A substantial body of literature supports the effectiveness of aided AAC interventions in promoting communication in individuals with communication disabilities (Ganz et al., 2012; Schlosser & Wendt, 2008).

[a]Department of Communication Sciences and Disorders, The Pennsylvania State University, University Park
[b]E. K. Shriver Center of the University of Massachusetts Medical School, Worcester

Correspondence to Jiali Liang: jiali.liang.aac@gmail.com

Editor-in-Chief: Sean Redmond
Editor: Geralyn Timler

Received September 3, 2017
Revision received November 12, 2017
Accepted December 20, 2017
https://doi.org/10.1044/2018_JSLHR-L-17-0331

Visual scene displays (VSDs) are one type of aided AAC layout. They present meaningful and naturalistic contexts such as a child's birthday party, sports game, and other outing/event (Blackstone, 2005). Concepts or messages associated with the event within these contexts could be programed into the displays (Wilkinson, Light, & Drager, 2012). These messages are stored via hotspots, in which selection of that hotspot results in activation of the messages contained within. For instance, touching the image of the birthday cake within a birthday party display may activate the programed message. The output can be, but is not limited to, a speech sound saying "birthday cake" or singing the birthday song or a written text "Happy Birthday!". Therefore, a VSD is a milieu through which individuals might communicate one or multiple messages (Wilkinson et al., 2012).

VSDs may hold promise as a layout option for aided AAC systems for a wide range of individuals. Drager, Light, Speltz, Fallon, and Jeffries (2003) have described the

Disclosure: The authors have declared that no competing interests existed at the time of publication.

proof-of-concept for using the contextually based scenes with young learners (Drager et al., 2003; Light et al., 2004). They noted that symbolic communication learning in young children with typical development (TD) benefits from a rich, contextually supported environment. VSDs are designed to provide a contextualized environment within which language concepts are depicted through rich events or routines familiar to the users. Empirical evidence has emerged to support that VSDs would be useful and beneficial to a wide variety of individuals with disabilities such as aphasia, Down syndrome (DS), and cerebral palsy (e.g., Hux, Buechter, Wallace, & Weissling, 2010; Light & Drager, 2008).

Display Design and Composition

A growing body of research concerning design of traditional grid-based displays has demonstrated that visual–perceptual characteristics of the display itself (symbol arrangement, color cues) can facilitate or inhibit the speed and accuracy of both visual search as well as reaching for a target symbol by individuals with and without developmental disabilities (Liang, Wilkinson, & Sainburg, 2018; Wilkinson, Carlin, & Thistle, 2008; Wilkinson & McIlvane, 2013; Wilkinson & Snell, 2011). Those studies suggest that the physical characteristics of grid displays can influence behaviors that are important to functional use (speed and accuracy of message preparation).

Logically, it seems quite likely that composition of VSDs may also positively or negatively influence visual attention to concepts within those displays (Wilkinson et al., 2012). A VSD that is poorly laid out or not well matched to the visual skills of its user may lead to inefficient visual attention to the meaningful elements in the display, which in turn would be expected to affect use during functional communication. Yet, there currently is little empirical evidence as to which elements and how various compositions of VSDs may contribute to visual attention to meaningful elements within VSDs (cf. Wilkinson et al., 2012). The evidence from studies of visual attention to traditional grids may not be generalizable to photographic displays, given the differences between these two types of layouts in terms of visual complexity, color, luminance, size, and social–perceptual features (cf. Wilkinson & Jagaroo, 2004). It is therefore critical to map out carefully, within displays that simulate VSDs, how the various visual characteristics of photographs affect individuals' visual attention to the elements within them.

Rationale for Studying Human Figures in VSDs

The rationale for use of VSDs reflects their ability to represent elements important to language development: (a) a familiar event context in which language concepts are embedded and (b) a person engaged in a socially related meaningful activity (Blackstone, 2005; Wilkinson et al., 2012). However, photographs typically also contain other elements in the background (furniture, pictures on the wall, etc.) that have the potential to distract attention from these meaningful elements (Wilkinson et al., 2012). Moreover, some evidence suggests that those with certain disabilities, such as autism spectrum disorder (ASD), might exhibit gaze avoidance to humans in photographs (e.g., Wilson, Brock, & Palermo, 2010). Clearly, it is necessary to examine whether composition of the VSDs influences gaze behavior toward the meaningful elements (human and meaningful activity).

A growing body of research has used automated eye-tracking technology to study visual attention to humans in photographs. Automated eye-tracking technologies enable recording of participants' point of gaze while they view visual stimuli and can offer insight into visual attention that could not be acquired via traditional methods (Wilkinson & Mitchell, 2014). In studies examining simulated VSDs, human figures have been found to attract visual attention within the first second of viewing a photograph and to maintain viewers' attention thereafter, both for children and adults with and without disabilities (Thiessen, Beukelman, Ullman, & Longenecker, 2014; Wilkinson & Light, 2011, 2014). This effect was present even when the human figures were small, placed far off the center of the scenes, and/or in competition with potentially distracting background items (Wilkinson & Light, 2011, 2014).

Although the presence of human figures seemed to exert strong visual attraction, it is also important to examine whether specific features of the human figures influence gaze. In a pioneering work by Yarbus (1967), typically developing participants preferentially looked at the faces of human figures (especially the eyes) in a scene, but this preference seemed to decline when the human figures alone were presented. Building on Yarbus's work, Birmingham, Bischof, and Kingstone (2008) found that fixations on the eye region increased as the number of people depicted and the level of activity between these human figures increased. The researchers proposed that, as more social content was depicted, increasing visual attention to the eyes helped to extract social information.

In the current study, we sought to examine two potential influences on visual processing of VSDs: (a) the number of human figures present and (b) the presence of sharing social activity between the human figures depicted. These two factors reflect the theoretical underpinnings of the designs of VSDs, as such displays are intended to support communication about socially meaningful events. However, if individuals who are provided with such VSDs do not attend to the human figures depicted within them, then a single drag-and-drop of human figures into the scenes may not make an effective visual scene for message representation. Examining this visual pattern in individuals who might use AAC, particularly those with ASD or DS, is also critical, as manufacturers and clinicians who construct and design such visual scenes must consider how these factors may support the use of the scenes for communication.

Effects of Humans in Scenes on Gaze Patterns of ASDs

One of the core symptoms of individuals with ASD is the presence of early and persistent deficits in social interaction and communication (American Psychiatric Association, 2013), including a lack of or alterations to normative patterns of eye contact during social interactions (Neumann, Spezio, Piven, & Adolphs, 2006). Some scholars have proposed that social–communicative deficits in ASD may result from atypical orienting and decreased attention to socially relevant stimuli such as faces in early life (Dawson, Webb, & McPartland, 2005; Elsabbagh et al., 2013; Schultz, 2005; Wass et al., 2015).

As reviewed by Guillon, Hadjikhani, Baduel, and Rogé (2014), a substantial body of eye-tracking studies has examined gaze behavior by individuals with ASD toward socially relevant stimuli (specifically human figures and faces). These studies have reported divergent results; some showed differences in visual attention between individuals with ASD and those with TD (Hanley, McPhillips, Mulhern, & Riby, 2013; Riby & Hancock, 2008; Wilson et al., 2010), but others reported that individuals with ASD, just as controls with TD, directed their gaze to faces in the first fixations (Elsabbagh et al., 2013; Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2009). Guillon et al. (2014) suggested that this may reflect diverse research designs. For example, Chevallier and colleagues (2015) examined visual attention to static versus dynamic stimuli, whereas Sasson, Turner-Brown, Holtzclaw, Lam, and Bodfish (2008) probed the sensitivity of circumscribed-interest-related objects versus common objects. Some researchers examined visual attention during free viewing versus after instruction (Fletcher-Watson et al., 2009; Fletcher-Watson, Leekam, Findlay, & Stanton, 2008). Equivocal findings may also be attributed to factors such as age, gender, IQ, autism symptomatology, or comparison group (cf. Rice, Moriuchi, Jones, & Klin, 2012).

Given the core difficulties in social interaction by those with ASD, it seems possible that the number of human figures and the presence of sharing activity between the human figures depicted in photographs might affect viewing behaviors. Many eye-tracking studies used one-person photographs, reporting no significant between-group differences in viewing patterns (e.g., Fletcher-Watson et al., 2009; Freeth, Ropar, Chapman, & Mitchell, 2010; Wilkinson & Light, 2014). In contrast, individuals with ASD spent less time viewing the faces than their peers with TD when the visual stimuli contained two or more persons (Riby & Hancock, 2008; Wilson et al., 2010). Therefore, one might question whether all human figures are equal. If individuals with ASD avoid viewing an increased number of human figures with the presence of sharing activity between the human figures, then VSDs with such elements might be suboptimal for this population.

Effects of Humans in Scenes on Gaze Patterns of Individuals With DS

DS is the most common chromosomal disorder with an estimated incidence of 1 in every 1,000 live births (Shin et al., 2009). Individuals with DS have particular deficits in expressive language and syntax as well as speech intelligibility (Roberts, Price, & Malkin, 2007). Receptive language may also be compromised (Chapman, 2006). AAC can therefore be used to promote the communication development of individuals with DS (Brady, 2008; Martin, Klusek, Estigarribia, & Roberts, 2009).

To our knowledge, no eye-tracking study has examined the two factors of interests (number of people and the presence of sharing activity) with individuals with DS. However, we might extrapolate from our understanding of the social profile in DS as well as one prior study of visual attention to photographs. The behavioral phenotype of DS includes strengths in daily living and socialization skills, and many individuals are reported as outgoing and social (e.g., Dykens, Hodapp, & Evans, 2006). In a preliminary study using eye-tracking technology with a very small sample, Wilkinson and Light (2014) reported that participants with DS fixated on human figures as rapidly and for as much time as vocabulary-matched children without disabilities. Both the reported behavioral phenotype and the one study of visual attention suggest that participants with DS might give preferential attention to humans in scenes, no matter how many figures are depicted.

Study Aims

The current study examined the following research question: What is the effect of the number of human figures (two vs. three) and the presence of sharing activity between the human figures (sharing-present vs. sharing-absent) in still photographs on the gaze patterns of individuals with ASD, DS, and TD? We anticipated that (a) the participants with TD would fixate on human figures within the sharing-present scene more than the sharing-absent scene and that, (b) as the number of human figures increased, more fixation time would be spent on the human figures. In contrast, it was possible that those with ASD might show fewer and slower fixations to human figures, in general, and also within the sharing-present scene than the sharing-absent scene. We anticipated that individuals with DS and peers with TD would respond faster to and fixate more than those with ASD on human figures within the sharing-present scene with an increased number of people depicted.

Method

Participants

Three groups of individuals participated: individuals with ASD, those with DS, and those with TD. Each group contained 10 individuals. Table 1 presents participants' gender, age, diagnostic status, and, for the groups with disabilities, standard scores on the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2006) test of receptive vocabulary. Individuals with ASD were matched with those with DS based on their receptive language (as described below) and chronological age. These two groups were chronological-age matched with the group with TD. The chronological age range was from 7 to 32 years (although most were older teenagers

Table 1. Participant characteristics.

Etiology	Participant code	Gender	CA	PPVT	CARS/GARS	Location
ASD	ASD1	M	19;3	71	84[a]	School
	ASD2	M	10;6	57	89[a]	School
	ASD3	M	17;0	N/T	104[a]	School
	ASD4	M	10;11	29	40[b]	School
	ASD5	M	16;4	20	43.5[b]	School
	ASD6	M	15;3	N/T	40[b]	School
	ASD7	M	16;2	58	NA	School
	ASD8	M	9;3	N/T	NA	Center
	ASD9	M	7;4	90	NA	Lab
	ASD10	M	18.9	20	37[b]	School
	Mean		14;1	49		
DS	DS1	F	21;4	74		Lab
	DS2	F	16;8	33		Lab
	DS3	F	23;0	55		Lab
	DS4	M	18;9	58		Lab
	DS5	M	31;11	43		Lab
	DS6	F	17;0	20		Lab
	DS7	M	11;2	39		School
	DS8	F	10;1	73		Lab
	DS9	M	12;8	52		Lab
	DS10	F	21;4	64		Lab
	Mean		19;0	51		
TD	TD1	F	10;3	All are functional at grade level or were enrolled in college.		Lab
	TD2	F	13;0			Lab
	TD3	M	15;2			Lab
	TD4	F	15;2			Lab
	TD5	F	21;8			Lab
	TD6	F	18;8			Lab
	TD7	M	22;1			Lab
	TD8	F	14;10			Lab
	TD9	F	23;1			Lab
	TD10	M	22;3			Lab
	Mean		17;8			

Note. CARS scoring: 15–29 = nonautistic, 30–36 = mildly–moderately autistic, and 37–60 = severely autistic. GARS-3 scoring: ≤ 54 = ASD unlikely; 55–70 = ASD probable; 71–100 = ASD very likely, requires substantial support; and ≥ 101 = ASD very likely, requires very substantial support. CA = chronological age (years;months); PPVT = standard score on the Peabody Picture Vocabulary Test–Fourth Edition; CARS = Childhood Autism Rating Scale; GARS = Gilliam Autism Rating Scale; Location = location of testing; ASD = autism spectrum disorder; M = male; School = testing occurred in a quiet room of the participant's school; N/T = not testable (i.e., the test was attempted, but the participant showed low compliance); NA = data unavailable at the point of reporting; Center = testing occurred in a quiet room of a therapy center attended by the participant; Lab = testing occurred in the research investigator's laboratory; DS = Down syndrome; F = female; TD = typical development.

[a]Participant diagnosis confirmation obtained via GARS. [b]Participant diagnosis confirmation obtained via CARS.

and adults; see Table 1). Statistical analysis of whether age influenced the pattern of results was conducted to address this potential issue. In addition, although these represent small sample numbers, the individual data for each participant are quite sizable, as 60 samples of gaze were obtained each second, for each of the 32 stimuli (described below), which were each presented for 5 s (hence, up to 9,600 samples per participant). The small sample size is considered in the Discussion.

A trained research assistant completed the PPVT administration. The scores of five participants in each of the clinical groups (DS and ASD) reflected moderate to severe vocabulary delay; two participants in each group had scores in the range of mild delay. We used the PPVT-4—a language measure—for matching individuals with ASD and individuals with DS rather than general or nonverbal intellectual functioning as the inclusion criterion, for two reasons. First, as in Wilkinson and Light (2014), our research targets the design of AAC, which is intended to support communication and language outcomes. Second, our study compared two very different clinical populations. Individuals with ASD often have lower verbal than nonverbal outcomes. Charman (2004) argued that full-scale or nonverbal IQ matching may selectively disadvantage their language. For example, many eye-tracking studies that used a nonverbal-intelligence-based inclusion criterion found between-group differences in viewing patterns. The receptive language scores were not statistically different between the two clinical groups, $p = .798$. There was no significant difference in chronological ages across all groups, $p = .163$.

Diagnosis of DS or autism was initially provided by parent report on a demographic questionnaire. Parent-reported diagnosis of autism was then confirmed via convergence of evidence. For 7 of the 10 participants with ASD, confirmation of diagnosis was indicated by their school placement (attending specialized school for individuals with ASD) and a score indicating presence of autism on either the Childhood Autism Rating Scale (CARS; Schopler, Reichler, & Renner, 1988) or Gilliam Autism Rating Scale (GARS [Third Edition]; Gilliam, 2014). Both the CARS and the GARS are checklists filled by parents, teachers, or other persons familiar with the individual. For three of the participants with ASD, no CARS/GARS score was available. Confirmation of diagnosis for these participants was provided by the parent report of diagnosis resulting from clinical assessment by a developmental pediatrician at a large center for autism (ASD8 and ASD9) and/or attendance at a specialized school (ASD7 and ASD9).

Individuals with TD were recruited through word of mouth. The remaining individuals with disabilities were recruited either through word of mouth or from two self-contained schools serving individuals with severe disabilities. All participants with disabilities had some experience with traditional row-column AAC systems, but they varied in the actual usage patterns. Although the participants were familiar with image-based media in general, none of them had exposure to systems using VSDs as defined in the AAC literature (Blackstone, 2005). Participants had vision within or corrected to within normal limits.

Materials: Photographs of Humans in Naturalistic Scenes

Sixteen pairs of photographs (32 in total) were purchased from a stock photography company (http://www.fotosearch.com/). Thirteen of the pairs had the same actors and settings/backgrounds depicted within them. The remaining three pairs had actors of similar ages and settings consistent with each other (e.g., young adults having lunch or dinner together, couple reading newspapers during breakfast). Figure 1 presents four exemplar photographs. The pairing of the photographs allowed the first experimental manipulation; in one photograph of each pair, there was a sharing activity between the human figures (sharing-present scene), whereas in the other photograph of that pair, the people were doing individual activities and sharing activity was absent (sharing-absent scene). Thus, for instance, in Panels C and D of Figure 1, the three individuals are either attending to their own individual devices (Panel C) or are instead all attending to the same device (Panel D). In addition, the photographs either contained two (Panels A and B) or three (Panels C and D) individuals. The use of stock photographs (rather than photographs personalized to each participant) was necessary to ensure close control of the variables of interest in this study; thus, like Wilkinson and Light (2014), these were simulations of VSDs but not necessarily ones that would be implemented clinically. Twelve filler photographs of minerals, plant sculptures, and landscapes and six video clips (animations) were intermixed with the target stimuli to ensure variety among the stimuli (e.g., for the mitigation of the perception of repetition).

Photographs were comparable on the size of the people within them, as quantified by the number of pixels they occupied. A two-way analysis of variance (ANOVA; sharing-present vs. sharing-absent, two-person vs. three-person) confirmed that there was no significant difference in size between sharing-present and sharing-absent humans depicted in the photographs, $p = .54$, and between two and three persons depicted, $p = .67$, and no interaction of the two variables, $p = .61$.

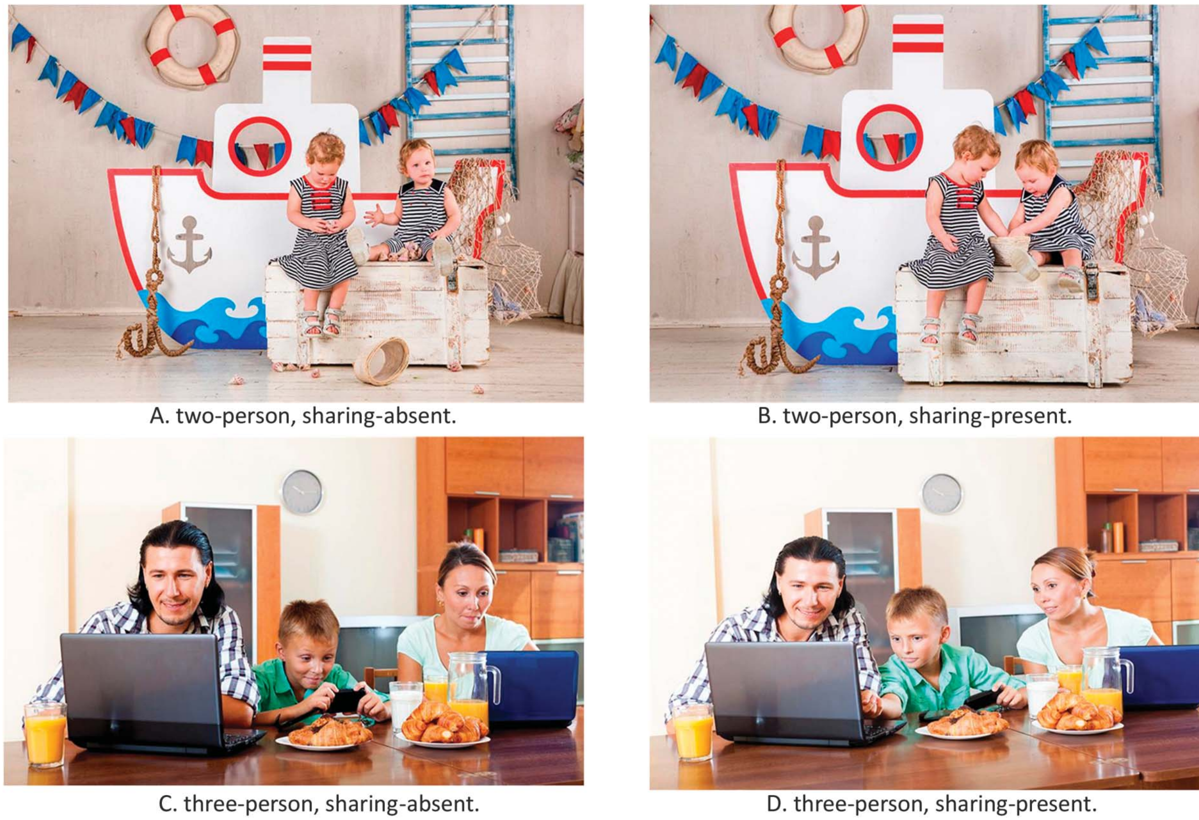
Data Acquisition and Eye-Tracking Equipment

All sessions occurred in a quiet room at the participant's school or the second author's research laboratory. Data were collected via Tobii T60 eye-tracking system. The 17-in. T60 monitor contained a built-in projection strip, by which infrared light was projected onto the participants' pupils and corneas. The light reflected off the eyes was captured by three built-in cameras. Point of gaze was recorded by the Tobii Studio software at a rate of 60 samples per second. A fixation was defined as a series of consecutive samples in which the gaze remained within a 35-pixel area for greater than 100 ms. Patterns of fixation on photographs were evaluated for each participant.

At the outset of each session, a calibration was performed before the presentation of stimuli to ensure accurate data acquisition. Participants sat approximately 65 cm from the screen. Two trained research assistants used the reading displayed on the monitor to adjust the location of the monitor or participant's seating to ensure that the standard was met. During the calibration, a brief video appeared first in the upper left corner and then in the lower right corner of the monitor. Accurate calibration was achieved when the participant's fixation to the video stimulus occurred within the two predefined points (upper left and lower right) within the display screen. The motion-based video was used to maximize the likelihood of attracting a fixation without the need for verbal instruction (Wilkinson & Light, 2014).

After calibration, participants were told, "You are going to see a series of pictures; just look at them as you would normally." Each photograph was then presented for 5 s. An interstimulus interval slide was presented for 1 s. This slide was blank (white) with a red fixation circle at the top. A free-viewing paradigm was considered to be particularly well suited to our research questions, as measurement of spontaneous behavior is a means to characterize visual patterns as they occur in the real world (Ames & Fletcher-Watson, 2010; Karatekin, 2007; Smilek, Birmingham, Cameron, Bischof, & Kingstone, 2006). This paradigm has been used in studies with individuals of a wide age range and etiologies such as ASD, DS, and Williams syndrome (e.g., Elsabbagh et al., 2013; Fletcher-Watson et al., 2009; Norbury et al., 2009; Riby & Hancock, 2008, 2009; Wilkinson & Light, 2014). We chose a laboratory-based free-viewing paradigm to map out the fundamental impact of the two factors of interest on visual attention in our three groups (Wilkinson & Light, 2011). It was necessary to examine visual attention in a nonsocial setting before

Figure 1. Examples of four of the photographs displayed. Images supplied from Fotosearch: (a) k17808031, (b) k17807745, (c) k19308684, and (d) k21944420.



any further study of use within actual communication interaction, given the different social phenotypes of our two clinical groups. The information obtained about visual attention in the absence of social demands can be examined in the future with regard to how attention patterns are (or are not) altered when social demands of actual interaction are introduced.

We used the Tobii Studio software to draw three areas of interest for each target photograph: the whole scene, the whole human figure (face and body), and the nonhuman area within each scene. The software matched the point-of-gaze coordinates obtained from the T60 monitor to the coordinates defined within each area of interest. This enables the software to calculate the fixation parameters defined as dependent measures.

Dependent Variables

The Tobii Studio software yields a set of fixation data to each area of interest in a chronological order of occurrence and with a time stamp. Consistent with Wilkinson and Light's (2011, 2014) research rationales and inquiry processes, we were interested in six measures: (a) the average total amount of time each participant spent fixating on the whole image during the 5-s display time, (b) the percentage of each participant's total fixation that was spent on any human figure, (c) the ratio of time spent fixating on the human figures in relation to its size, (d) the ratio of time spent fixating on the head in relation to its size, (e) each participant's average absolute latency to produce the first fixation on the human figures across the 32 photographs (absolute latency was defined as the amount of time elapsed between the onset of image presentation on the screen and the start of the first fixation), and (f) each participant's average relative latency to produce the first fixation on the human figures relative to the participant's own first fixation on the screen. A seventh measure, percentage of photographs in which participants fixated on the human figures, was also included to examine how likely that a photograph was to attract a participant's attention at all when a photograph contained human figures (see Supplemental Material S1). Operational definitions for each measure and rationale for using these measures are summarized in Supplemental Material S1.

Data Analysis

Mixed ANOVA allowed evaluation of the role of the presence of sharing activity (sharing-present vs. sharing-absent; within subjects), the number of people (two- vs.

three-person; within subjects), and group (TD vs. ASD vs. DS; between subjects) or their interaction on the dependent measures. Because of the wide variation in chronological age, a preliminary analysis of covariance was conducted in which chronological age was added as a covariate to assess if age influenced (alone or with other variables) any of the outcomes. Age was never a significant covariate for any of the variables, and there were no any significant interactions between age and the other measures. We therefore removed the age variable from the equation to preserve degrees of freedom/power, and all analyses were conducted with only the main variables of interest. The parametric analysis was carried out as an omnibus starting point, with inspections of any violation of assumptions. Post hoc testing was conducted via both parametric and nonparametric analyses to confirm that the results were not artifacts of the small sample size.

Results

Mean Total Time Spent Fixating on the Photograph Across All Elements

This measure reflects the total amount of the 5-s period in which fixations were recorded. It was calculated by summing all the fixations anywhere within the visual stimuli. If participants blinked, looked away, or made head movements that caused temporary loss of the eye image, the eye tracker would register the period as the time not spent fixated. Mixed 2 × 2 × 3 ANOVA showed a significant main effect of the number of people depicted, $F(1, 27) = 4.459$, $p = .044$, partial $\eta^2 = .142$, where participants seemed to fixate on the three-person photographs for a longer time than the two-person ones (Figure 2). A significant group difference was identified, $F(2, 27) = 28.491$, $p < .001$, partial $\eta^2 = .679$. Games–Howell post hoc procedure was used because the homogeneity of variance assumption was not met. The group with ASD fixated significantly less than the group with DS, $p = .035$. Both the group with ASD, $p < .001$, and the group with DS, $p < .001$, fixated significantly less than the group with TD. There were no main effects of the presence of sharing activity or two- or three-way interactions.

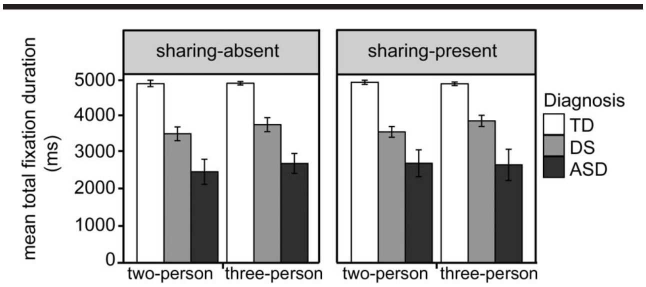


Figure 2. Mean total time spent fixating on the photograph across all elements. TD = typical development; DS = Down syndrome; ASD = autism spectrum disorder.

The effect sizes partial η^2 for these nonsignificant results were minimal to very small (range = .009–.150).

Mean Percentage of Each Participant's Own Time Spent Fixating on the Human

As each group differed significantly in their total amount of time spent on the photographs, it is important to use each participant's own time for the percentage measure rather than the fixed 5-s photograph display time. The mean percentage of a participant's own fixation time that was spent fixating on the human figures was calculated by dividing the participant's viewing time on human figures by his or her own viewing time on the screen (Figure 3). Mixed 2 × 2 × 3 ANOVA indicated a significant main effect of the number of people depicted, $F(1, 27) = 11.086$, $p = .003$, partial $\eta^2 = .291$, where there was a higher percentage of time spent on the humans within the two-person than the three-person photographs. However, there was no significant effect of the presence of sharing activity, $p = .21$; no between-group differences, $p = .184$; and no significant two- or three-way interactions. The effect sizes partial η^2 for all nonsignificant results were small (range = .022–.129).

Ratio of Time Spent Fixating on the Human Figures Relative to Their Size

This ratio measure was calculated by dividing the actual time spent on the human figures by the expected amount of time spent the human figures based on their size (Figure 4). Ratios over 1 mean that the participants spent more time fixated on the human figures than would be expected based on the human figures' size. Mean ratios were greater than 1.5 across all photograph types and groups, demonstrating that the human figures in the scenes attracted more attention than would be expected, based on their size alone. Mixed 2 × 2 × 3 ANOVA indicated a significant main effect of the number of people depicted, $F(1, 27) = 10.032$, $p = .004$, partial $\eta^2 = .271$, where there was a higher ratio of time spent on the humans within the two-person than the three-person photographs. There was also a significant interaction between the diagnosis and the number of people depicted, $F(2, 27) = 4.918$, $p = .015$, partial $\eta^2 = .267$.

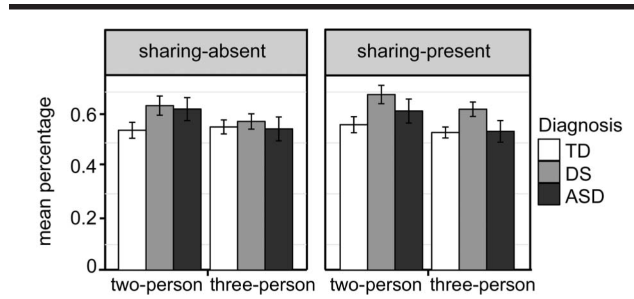


Figure 3. Mean percentage of each participant's own time spent fixating on the human figures. TD = typical development; DS = Down syndrome; ASD = autism spectrum disorder.

Figure 4. Ratio of time spent fixated on the human figures relative to their size. TD = typical development; DS = Down syndrome; ASD = autism spectrum disorder.

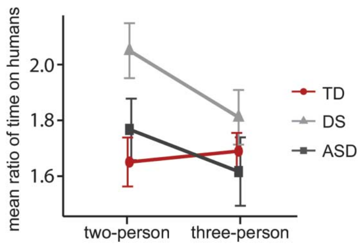
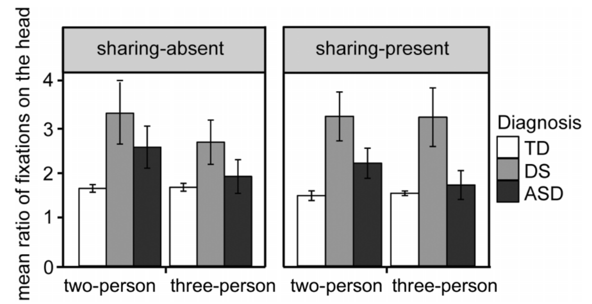


Figure 5. Ratio of time spent fixated on the head relative to its size. TD = typical development; DS = Down syndrome; ASD = autism spectrum disorder.



Post hoc paired-sample *t* tests revealed that only the group with DS exhibited a significantly higher ratio of time within the two-person photos than the three-person ones, $t(9) = 3.164$, $p = .011$. There was a significant main effect of the presence of sharing activity, $F(2, 27) = 10.481$, $p = .003$, partial $\eta^2 = .280$, where there was a higher ratio of time spent on the humans within the sharing-absent than the sharing-present photographs. No group differences, $p = .118$, or two- or three-way interactions were detected. The effect sizes partial η^2 for all nonsignificant results were very small (range = .022–.145).

One sample *t* tests evaluated if the calculated ratio was significantly different from 1, for each group. The *p* value was initially set to .05 and then divided by 4 (the number of conditions) to adjust for multiple comparisons; thus, the final *p* value criterion for judging whether a ratio was different from 1 was $p = .0125$. Table 2 suggests that the ratio was determined to be of statistical significance at this adjusted *p* value in all types of photographs, regardless of etiology.

Ratio of Time Spent Fixating on the Head Relative to Its Size

The next analysis focused on fixations on the head or face of the persons in the photographs, rather than the whole person. This ratio measure was calculated by dividing the actual time spent on the head by the expected amount of time spent on the head based on its size. As Figure 5 illustrates, the ratio approached or exceeded 2 across groups, which means that all participants fixated on the heads more than expected based on their sizes alone. Mixed 2 × 2 × 3 ANOVA revealed a significant main effect of the number of people depicted, $F(1, 27) = 7.064$, $p = .013$, partial $\eta^2 = .214$, where the heads were fixated on more in the two-person photographs than the three-person ones. There was a significant group difference, $F(2, 26) = 4.247$, $p = .025$, partial $\eta^2 = .246$. Games–Howell post hoc procedure showed that the ratio was higher in the group with DS than the group with TD, $p = .049$. However, there was no other between-group difference. There was no main effect of the presence of sharing activity, $p = .967$, and no two- or three-way interaction effect. These nonsignificant results had minimal to small effect sizes (range = .000–.167).

Mean Absolute Latency of Fixation on the Human Figures

Mean absolute latency refers to the time elapsed between the onset of the presentation of the image and the first fixation to the human figures. Figure 6 shows that the group with TD launched their first fixations on the human figures within 0.5 s across photographs ($M = 389.05$ ms, $SD = 591.30$ ms). All participants with DS first fixated on the human figures within 1 s ($M = 606.57$, $SD = 780.50$). In contrast, the group with ASD was slower than the other

Table 2. Summary of *t* test analyses of the ratio measure for each stimulus type.

	TD			ASD			DS		
Stimulus type	**Mean (*SD*)**	***t* value**	**95% CI**	**Mean (*SD*)**	***t* value**	**95% CI**	**Mean (SD)**	***t* value**	**95% CI**
Two-person, sharing-absent	1.72 (0.27)	8.49**	[0.53, 0.91]	1.93 (0.47)	6.32**	[0.60, 1.26]	2.11 (0.47)	7.48**	[0.77, 1.44]
Two-person, sharing-present	1.57 (0.34)	5.22*	[0.32, 0.81]	1.61 (0.32)	6.08**	[0.38, 0.84]	1.98 (0.27)	11.58**	[0.79, 1.18]
Three-person, sharing-absent	1.75 (0.28)	8.58**	[0.55, 0.95]	1.74 (0.44)	5.35**	[0.43, 1.05]	1.78 (0.32)	7.63**	[0.55, 1.01]
Three-person, sharing-present	1.62 (0.16)	12.29**	[0.50, 0.74]	1.49 (0.45)	3.41*	[0.16, 0.81]	1.83 (0.35)	7.44**	[0.58, 1.08]

Note. $N = 10$ participants for each group. TD = typical development; DS = Down syndrome; ASD = autism spectrum disorder; CI = confidence interval.

$^*p < .01$. $^{**}p < .001$.

Figure 6. Mean absolute latency of fixation to the human figures. TD = typical development; DS = Down syndrome; ASD = autism spectrum disorder.

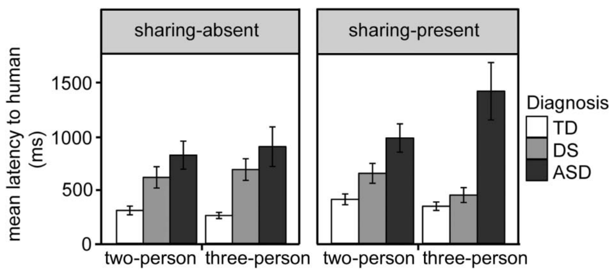
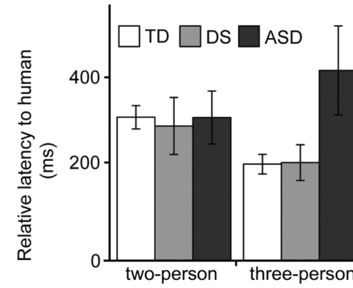


Figure 7. Mean relative latency of fixation to the human figures. TD = typical development; DS = Down syndrome; ASD = autism spectrum disorder.



two groups to first fixate on the human figures (M = 1003.12, SD = 1142.85).

Mixed 2 × 2 × 3 ANOVA confirmed a significant group difference, $F(2, 27) = 18.98$, $p < .001$, partial $\eta^2 = .584$. Games–Howell post hoc procedure revealed that the group with ASD fixated on the human figures significantly slower than the group with TD, $p = .001$, and the group with DS, $p = .022$. The group with DS was also significantly slower than the group with TD, $p = .003$. There was a significant interaction between the diagnosis and the number of people depicted, $F(2, 27) = 5.922$, $p = .007$, partial $\eta^2 = .305$. Related sample Wilcoxon signed rank tests indicated that the group with TD was significantly slower to first fixate on the human figures within the two-person photographs than the three-person photographs, $Z = -2.090$, $p = .037$. In contrast, the group with ASD was significantly slower to first fixate on the human figures within the three-person photographs than the two-person ones, $Z = -1.988$, $p = .047$. The group with DS did not show any difference for the two types of photographs, $p = .33$. There was no effect of the number of people depicted, $p = .302$; the effect of the presence of sharing activity, $p = .153$; or other interaction effects. The effect sizes for the nonsignificant results were very small (range = .001–.166).

Mean Relative Latency of Fixation on the Human Figures

Mean relative latency refers to the amount of time elapsed between each participant's first fixation anywhere on the image and the first fixation on the human figures. Figure 7 illustrates that, regardless of etiological status, most participants first fixated on the humans within 500 ms across photographs. Mixed 2 × 2 × 3 ANOVA revealed a significant interaction between the diagnosis and the number of people depicted, $F(2, 27) = 3.574$, $p = .042$, partial $\eta^2 = .209$. Post hoc one-way ANOVAs revealed that the group with TD was significantly slower to first fixate on humans in two-person photographs than three-person ones, $F(1, 9) = 32.067$, $p < .001$, partial $\eta^2 = .781$. However, the number of people did not have an effect on the group with DS, $p = .174$, or the group with ASD, $p = .239$. There were no main effects of the presence of sharing activity, $p = .170$; the number of people depicted, $p = .367$; or group differences, $p = .265$, or any other interactions. All effect sizes for the nonsignificant results were minimal to small (range = .004–.113).

Discussion

Overall gaze behaviors were quite similar across groups, corroborating not only the proposal that human figures should be added (Gillespie-Smith & Fletcher-Watson, 2014; Wilkinson & Light, 2014) but also the idea that the presence of social content (people engaged in sharing activities) attracts attention in all groups. However, significant differences emerged between the three groups in the percentage of photographs in which humans were fixated (see Supplemental Material S1), the total amount of screen viewing time, and the absolute latency to first view the human figures. These findings have implications for visual response to photographs containing social content across the three groups as well as the ways in which AAC design could respond to such visual patterns.

Social Content Attracts Visual Attention

Social content has been proposed to be a critical component for visual attention to social elements (Birmingham et al., 2008; Guillon et al., 2014). Birmingham et al. (2008) speculated that, for a one-person scene, actions draw attention away from the eyes because eye information is not critical for understanding; in contrast, when the number of people and the level of social activity increased, the eyes offer critical information. Those authors found that more people added to the photograph increased the extent that the eyes were scanned, but only when the people in the scene were actively doing something. We indexed the social contents via the number of people and the presence of sharing activity. Increasing the number of people depicted increased the total amount of time spent on the whole screen, in all groups. Increasing the number of people resulted in a faster launch of fixation on the human figures. However, this latter pattern only applied to individuals with TD. The presence of sharing activity seemed to decrease the ratio of time spent on the human relative to its size.

Including Multiple Clinical Groups Differentiates Syndrome-Specific Behavior

When calculated as an absolute latency (time from presentation of the stimulus to fixation on the human), individuals with ASD were slower to first fixate on the human figures as the number of people depicted increased, whereas those with DS seemed to be equally fast within two- versus three-person scenes. However, this effect of the number of people depicted disappeared in both groups when we calculated latency to fixate on the human figures relative to each participant's own overall speed to fixate. In other words, once participants started to fixate anywhere on the screen, the latency to make a fixation on the human figures was no different between participants with ASD and those with DS. This finding suggests that, although individuals with ASD may take longer to fixate on human figures than individuals with DS, their speed to fixate on human figures once an initial fixation has been established is no different.

In contrast, the group with TD prioritized visual attentions to human figures faster within the three-person photos than the two-person ones in both absolute and relative latency measures. The group with TD seemed to be more sensitive to the two- versus three-figure comparisons than the groups with ASD and DS. Future studies should explore this finding further, particularly to evaluate the possibility that those with ASD and DS take a longer time to process the visual information. Potentially valuable avenues for future research include extending the display time (see discussion in the next section) and adding more human figures. For example, VSDs with two versus five or six figures might yield different results.

The inclusion of participants in two clinical groups, all of whom had vocabulary limitations associated with intellectual/developmental disability, allowed evaluation of whether any deviation of any visual attention from those of typically developing individuals is autism specific. Compared with the group with TD, both individuals with ASD and those with DS viewed a lower percentage of photographs, spent less time viewing the screen, and were slower to initiate their first fixations on the human figures after the onset of the stimulus. These results emphasized the importance of adjusting our following analyses by considering participants' own viewing time and latency. Once the adjustment was made, two important findings were revealed: (a) The group with ASD resembled the group with TD across outcome measures except the relative latency, and (b) on the relative latency measure, the group with ASD and the group with DS were comparable with one another and different from the group with TD. This suggested that, in eye-tracking studies of ASD, it is necessary to include a control group of individuals with other developmental disabilities to avoid biased conclusions regarding atypicality in gaze behaviors in autism.

Inconsistent Findings May Reflect Study Design

Increasing the number of people depicted decreased the likelihood of visual scanning of the whole human figures. This is reflected in the following measurement: the percentage of one's own viewing time on the human figures, the ratio of time spent on the whole human figures, and the heads alone relative to their respective size. In addition, the presence of sharing activity also decreased the ratio of time spent on the human figure relative to their size. This finding is different from earlier reports (e.g., Birmingham et al., 2008; Yarbus, 1967). Several possibilities could explain these different outcomes. For instance, Birmingham et al. (2008) revealed that the impact of social content did not first emerge until after 6 s of viewing time. The current study only displayed each photograph for 5 s. Given more display time, participants might preferentially view sharing-present three-person scenes over sharing-absent two-person scenes. It is important to note, however, that the social cues or communicative messages, especially through the face, are often fleeting. We utilized a brief (5 s) display time to capture a more ecologically valid representation of how individuals react to these visual stimuli. If inclusion of social content such as people interacting does not provoke gaze avoidance in those with ASD, photographic VSDs can be potentially used to capture and represent the fleeting social communicative message. Additional instructional scaffolding from the partner may also help to encourage a better understanding of social information (Shane, 2006).

Second, Birmingham et al. (2008) examined the visual attention to the eyes of the people in the scene, whereas we evaluated the whole body and the head alone. Yu and Smith (2016) suggested that not only eyes but also other body parts (e.g., hands) are important sensory–motor pathways that young communicators rely on for social interaction and communication. Thus, when we focused on the head, there was either no difference (TD, ASD) or a higher (DS) ratio of time spent on the heads within the two-person than the three-person photographs. Systematic examination is needed to evaluate how gaze patterns toward eyes alone would be different, if any, from those toward the head alone and/or the body as a whole, for both the groups with TD and ASD and/or the group with DS.

The current study adds to a growing number of studies that were conducted with school-aged participants or young adults with ASD (Fletcher-Watson et al., 2009; Kemner, van der Geest, Verbaten, & van Engeland, 2007; Norbury et al., 2009; van der Geest, Kemner, Verbaten, & van Engeland, 2002; Wilkinson & Light, 2014). With the exception of absolute latency, individuals with ASD exhibited few differences in fixation patterns to scenes of various social contents. On the surface, this is inconsistent with the socio-communicative profile of autism (e.g., altered eye contact with people); this also appears to differ from research findings that individuals with ASD exhibited atypical or attenuated attention allocation to socially salient aspects of the scenes (e.g., Hanley et al., 2013; Riby & Hancock, 2008; Wilson et al., 2010). However, as noted by many researchers (e.g., Guillon et al., 2014; Rice et al., 2012; Wilkinson & Light, 2014), the different outcomes might be attributed to variabilities in the research designs and participant profiles. For instance, the individuals in the current study were school-aged children and young adults at the outset of diagnosis. They have likely been receiving at least some form of support or intervention services for most of their lives (Wilkinson & Light, 2014). It is possible that their visual attention has been positively influenced by intervention and/or maturation.

A strength of the current study is that we employed still photographic stimulus sets that were carefully controlled and analyzed. The social content as indexed by the number of people and the presence of sharing activity guided the selection procedures. For example, the actors, their sizes, and background objects depicted across stimuli were manipulated to maximize their comparability. In the data analysis, the sizes of the social element were also considered to rule out the possible confounding factor that typical attention to human figures was simply proportional to their sizes.

Attentional Patterns May Help to Inform VSD Design

The general cross-group similarities of gaze patterns in viewing photographs of various social contents suggest several things. Photographs including social content did not provoke gaze avoidance. Rather, they generally exerted strong attraction for visual attention, both for individuals with TD and those with ASD or DS, at least for the first 5 s. These findings underscored that the social contents constructed by two or more human figures are powerful draws on visual attention.

The default setting of dragging and dropping learning elements (e.g., pillow in the bedroom) by manufacturers may not be sufficient to support social communication (Wilkinson & Light, 2014). The design of VSD is built on the rationale that language learning occurs and is supported in a meaningful and motivating event (e.g., Wilkinson & Light, 2014). Human figures in photographs were visually attractive, which showed the potential of creating a naturalistic social scene for language learning (Wilkinson & Light, 2014). For example, they may be selected as hotspots for storing messages or key points for discussion (Thiessen et al., 2014).

During the 5-s viewing period, the number of human figures seemed to affect people's visual attention. Compared with three-person photographs, participants had a higher percentage of their own time spent fixating on human figures and a higher ratio of time spent on fixating on human figures as a whole or the heads alone within the two-person photographs. At least two possibilities may explain this pattern. First, it might take a longer time to process the complex social information within three-person scenes than two-person ones. Individuals were found to fixate on eyes more with the increasing number of people, suggesting that the complex level of scene analysis may not occur initially but may develop after a period of exposure time (Birmingham et al., 2008). If this is true, clinicians presenting VSD should ensure that individuals have a sufficient amount of time to inspect the scenes, especially when the number of people increases. Another possibility is that the gaze cues are more directly and explicitly presented via eye gaze, head direction, facial expressions, or body posture by three human figures such that participants may be readily driven to shift their attention to the nonhuman elements (objects or background) to infer the attentional states of the people depicted and the overall meaning of the scene. If this is true, VSD that contains more human figures might have the ability to guide viewers' attention to regions of the displays that might otherwise go without fixations (Thiessen et al., 2014). Analysis of gaze sequences regarding where and how participants moved from one element to the next might help to test the two possibilities.

Another clinical implication concerns the application of VSDs across populations. The group differences in the absolute latency measure were not observed in the relative latency to first fixate on the human figures. Individuals with ASD may have more difficulties to initiate the processing of visual stimulus rather than that of social elements. This is consistent with the argument that between-group differences identified in previous studies might reflect differences in total time, not relative time. Future research should examine what contributes to the prolonged time to initiate first fixations on the whole stimulus. If this reflects deficits in orienting to visual contents in general within both individuals with ASD and those with DS, careful design and use of photographic VSDs are needed. First, although images of familiar people and locations are readily obtained or generated with digital photographs, clinicians should be careful about selecting various compositions within VSDs so as not to overwhelm participants with unnecessary and distracting visual elements that are irrelevant to communicative tasks. Second, in light of delayed fixation to social content and to the whole screen in general for both individuals with ASD and DS, the display should be available for a sufficient time so that participants can first react to the display. Structured prompts or instructions may be needed to strategically guide user's attention. If necessary, visual scenes may be unfolded or presented one bit at a time rather than having full exposure of the whole scene all at once.

Limitations and Future Research Directions

The small sample sizes limit the generality of the findings. Nonparametric analyses were used where assumptions of parametric assumptions were violated to minimize any artifact results, and the capacity of eye-tracking technology to sample participants' gaze profiles at a high frequency may also help to collect representative data. Replication with a larger sample size will provide additional confidence that the nonparametric analyses reflect valid outcomes. For example, chronological age was not associated with any of the measures of interest in the current study. In a meta-analysis of research evidence from eye-tracking studies, chronological age of individuals with ASD was positively correlated with the resemblance between this group and the individuals with TD in terms of overall screen viewing time (Liang & Wilkinson, submitted). Examination of individuals with a wide age range may help to delineate the age effects on other measures of interest.

A fuller characterization of the sample would allow a more detailed discussion as it relates to the design of VSDs. For example, communication functioning, repetitive behaviors, visual acuity, autism symptomatology, and other potential variables might contribute to the data quality and results (e.g., Snell et al., 2010). Knowing participant characteristics such as social communication could help to examine the relationship between gaze patterns in a control setting and functional behavior in a natural environment (Wilkinson & Light, 2014).

Observing the image does not guarantee the processing of the information. Although there is evidence that location of visual fixation is strongly associated with attention (Rayner, 2009), further studies are warranted to distinguish between the presence of a simple fixation (observing) and the actual processing of the visual information (attending) within the VSDs (Dube et al., 2010; Dube & Wilkinson, 2014; Wilkinson & Light, 2014). The question of how fixation duration is associated with actual processing has not been answered conclusively. Durations of some fixations are determined irrespective of scene presence, whereas others seem to be correlated with "moment-to-moment control of scene analysis" (Rayner, 2009, p. 1481). Limited fixation on human figures in a VSD is not a prerequisite for use of VSD. It does not necessarily inhibit the processing of the information and preclude the ability to learn to use such displays as a communication symbol. Therefore, a critical next step would be to determine how we could best support the comprehension/actual processing of the various social contents in the photographs. A possible avenue would be to couple an eye-tracking passive viewing paradigm with behavior and/or verbal response within authentic communicative environments. For example, introducing tasks with instructions might help to differentiate what the group with ASD views spontaneously and what they could do when instructed (e.g., Fletcher-Watson et al., 2009). We could also examine visual scanning of the human eyes rather than the whole body or the head only because fixations on the eyes might be a subtle yet powerful way to examine social understandings in individuals with ASD (Birmingham et al., 2008).

To maintain consistency, we used the same set of stimuli for all participants. The photographs were highly controlled such that only the number of people and the presence of sharing activity varied. This rigorous control made it impossible to incorporate individualized stimuli that could have been more meaningful or motivating to the participants. Personalized materials are recommended for use in AAC systems (Thiessen et al., 2014). Thus, systematic comparisons are needed to investigate how visual scenes tailored to each participant might differ from generic scenes like the ones we used.

Conclusion

Individuals with ASD were similar to those with DS when they viewed naturalistic photographs that depicted highly complex social activities between human figures. When each participant's own pace and viewing time were considered, the two groups largely resembled typically developing individuals. Future studies should consider the selection of outcome measures and the inclusion of another clinical group to determine the generality or specificity of visual attention patterns in ASD. Clinicians who utilize visual scenes for message representation should consider users' viewing patterns (e.g., a general slower pace for the two clinical groups in this study) to maximize the effectiveness of the scenes they select. The current study provides an avenue of understanding compositions within ASD that are of potential interest to users as a way to improve the design of AAC systems.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, fifth edition: DSM-V*. Washington, DC: Author.

Ames, C., & Fletcher-Watson, S. (2010). A review of methods in the study of attention in autism. *Developmental Review, 30*(1), 52–73.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *The Quarterly Journal of Experimental Psychology, 61*(7), 986–998.

Blackstone, S. (2005). What are visual scene displays? *RERC on Communication Enhancement, 1,* 2.

Brady, N. C. (2008). Augmentative and alternative communication for children with Down syndrome or fragile X syndrome. In J. Roberts, R. Chapman, & S. F. Warren (Eds.), *Speech and language development and intervention in Down syndrome and fragile X syndrome* (pp. 255–274). Baltimore, MD: Paul H. Brookes.

Chapman, R. (2006). Language learning in Down syndrome: The speech and language profile compared to adolescents with cognitive impairment of unknown origin. *Down Syndrome Research and Practice, 10*(2), 61–66.

Charman, T. (2004). Matching preschool children with autism spectrum disorders and comparison children for language ability: Methodological challenges. *Journal of Autism and Developmental Disorders, 34*(1), 59–64.

Chevallier, C., Parish-Morris, J., McVey, A., Rump, K. M., Sasson, N. J., Herrington, J. D., & Schultz, R. T. (2015). Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Research, 8*(5), 620–628.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Journal of Developmental Neuropsychology, 27*(3), 403–424.

Drager, K. D. R., Light, J. C., Speltz, J. C., Fallon, K. A., & Jeffries, L. Z. (2003). The performance of typically developing 2 1/2-year-olds on dynamic display AAC technologies with different system layouts and language organizations. *Journal of Speech, Language, and Hearing Research, 46*(2), 298–312.

Dube, W. V., Dickson, C. A., Balsamo, L. M., O'Donnell, K. L., Tomanari, G. Y., Farren, K. M., & McIlvane, W. J. (2010). Observing behavior and atypically restricted stimulus control. *Journal of Experimental Analysis of Behavior, 94*(3), 297–313.

Dube, W. V., & Wilkinson, K. M. (2014). The potential influence of stimulus overselectivity in AAC: Information from eye tracking and behavioral studies of attention with individuals with intellectual disabilities. *Augmentative & Alternative Communication, 30*(2), 172–185.

Dunn, L. M., & Dunn, L. M. (2006). *Peabody Picture Vocabulary Test–Fourth Edition*. Circle Pines, MN: AGS.

Dykens, E. M., Hodapp, R. M., & Evans, D. W. (2006). Profiles and development of adaptive behavior in children with Down syndrome. *Down Syndrome Research and Practice, 9*(3), 45–50.

Elsabbagh, M., Gliga, T., Pickles, A., Hudry, K., Charman, T., & Johnson, M. H. (2013). The development of face orienting mechanisms in infants at-risk for autism. *Behavioural Brain Research, 251,* 147–154.

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia, 47*(1), 248–257.

Fletcher-Watson, S., Leekam, S. R., Findlay, J. M., & Stanton, E. C. (2008). Brief report: Young adults with autism spectrum disorder show normal attention to eye-gaze information-evidence from a new change blindness paradigm. *Journal of Autism and Developmental Disorders, 38*(9), 1785–1790.

Freeth, M., Ropar, D., Chapman, P., & Mitchell, P. (2010). The eye gaze direction of an observed person can bias perception, memory, and attention in adolescents with and without autism spectrum disorder. *Journal of Experimental Child Psychology, 105*(1), 20–37.

Ganz, J. B., Earles-Vollrath, T. L., Heath, A. K., Parker, R. I., Rispoli, M. J., & Duran, J. B. (2012). A meta-analysis of single case research studies on aided augmentative and alternative communication systems with individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*(1), 60–74.

Gillespie-Smith, K., & Fletcher-Watson, S. (2014). Designing AAC systems for children with autism: Evidence from eye tracking research. *Augmentative and Alternative Communication, 30*(2), 160–171.

Gilliam, J. E. (2014). *Gilliam Autism Rating Scale* (3rd ed.). Austin, TX: Pro-Ed.

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews, 42,* 279–297.

Hanley, M., McPhillips, M., Mulhern, G., & Riby, D. M. (2013). Spontaneous attention to faces in Asperger syndrome using ecologically valid static stimuli. *Autism, 17*(6), 754–761.

Hux, K., Buechter, M., Wallace, S., & Weissling, K. (2010). Using visual scene displays to create a shared communication space for a person with aphasia. *Aphasiology, 24*(5), 643–660.

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Cognitive Neuroscience, 27*(3), 283–348.

Kemner, C., van der Geest, J. N., Verbaten, M. N., & van Engeland, H. (2007). Effects of object complexity and type on the gaze behavior of children with pervasive developmental disorder. *Social Cognitive and Affective Neuroscience: Developmental and Clinical Perspectives, 65*(1), 107–111.

Liang, J., Wilkinson, K., & Sainburg, R. L. (2018). Is hand selection modulated by cognitive-perceptual load? *Neuroscience, 369,* 363–373.

Liang, J., & Wilkinson, K. M. (submitted). Gaze toward naturalistic social scenes by individuals with autism: A meta-analysis.

Light, J., & Drager, K. (2008). Evidence-based AAC interventions to build language and communication skills with infants, toddlers, and preschoolers. In Seminar presented at the ISAAC conference, Montreal, Canada.

Light, J., Drager, K., McCarthy, J., Mellott, S., Millar, D., Parrish, C., & Welliver, M. (2004). Performance of typically developing four- and five-year-old children with AAC systems using different language organization techniques. *Augmentative and Alternative Communication, 20*(2), 63–88.

Martin, G. E., Klusek, J., Estigarribia, B., & Roberts, J. E. (2009). Language characteristics of individuals with Down syndrome. *Topics in Language Disorders, 29*(2), 112–132.

Neumann, D., Spezio, M. L., Piven, J., & Adolphs, R. (2006). Looking you in the mouth: Abnormal gaze in autism resulting from impaired top–down modulation of visual attention. *Social Cognitive and Affective Neuroscience, 1*(3), 194–202.

Norbury, C. F., Brock, J., Cragg, L., Einav, S., Griffiths, H., & Nation, K. (2009). Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. *Journal of Child Psychology & Psychiatry, 50*(7), 834–842.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506.

Riby, D., & Hancock, P. J. B. (2008). Viewing it differently: Social scene perception in Williams syndrome and autism. *Neuropsychologia, 46*(11), 2855–2860.

Riby, D., & Hancock, P. J. B. (2009). Looking at movies and cartoons: Eye-tracking evidence from Williams syndrome and autism. *Journal of Intellectual Disability Research, 53*(2), 169–181.

Rice, K., Moriuchi, J. M., Jones, W., & Klin, A. (2012). Parsing heterogeneity in autism spectrum disorders: Visual scanning of dynamic social scenes in school-aged children. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(3), 238–248.

Roberts, J. E., Price, J., & Malkin, C. (2007). Language and communication development in Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(1), 26–35.

Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S., & Bodfish, J. W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Research, 1*(1), 31–42.

Schlosser, R. W., & Wendt, O. (2008). Effects of augmentative and alternative communication intervention on speech production in children with autism: A systematic review. *American Journal of Speech-Language Pathology, 17*(3), 212–230.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale*. Los Angeles, CA: Western Psychological Services.

Schultz, R. T. (2005). Developmental deficits in social perception in autism: The role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience, 23*(2), 125–141.

Shane, H. C. (2006). Using visual scene displays to improve communication and communication instruction in persons with autism spectrum disorders. *SIG 12 Perspectives on Augmentative and Alternative Communication, 15*(1), 8–13.

Shin, M., Besser, L. M., Kucik, J. E., Lu, C., Siffel, C., & Correa, A. (2009). Prevalence of Down syndrome among children

and adolescents in 10 regions of the United States. *Pediatrics, 124*(6), 1565–1571.
Smilek, D., Birmingham, E., Cameron, D., Bischof, W., & Kingstone, A. (2006). Cognitive etiology and exploring attention in real-world scenes. *Brain Research, 1080*(1), 101–119.
Snell, M. E., Brady, N., McLean, L., Ogletree, B. T., Siegel, E., Sylvester, L., & Sevcik, R. (2010). Twenty years of communication intervention research with individuals who have severe intellectual and developmental disabilities. *American Journal on Intellectual and Developmental Disabilities, 115*(5), 364–380.
Thiessen, A., Beukelman, D., Ullman, C., & Longenecker, M. (2014). Measurement of the visual attention patterns of people with aphasia: A preliminary investigation of two types of human engagement in photographic images. *Augmentative and Alternative Communication, 30*(2), 120–129.
van der Geest, J. N., Kemner, C., Verbaten, M. N., & van Engeland, H. (2002). Gaze behavior of children with pervasive developmental disorder toward human faces: A fixation time study. *The Journal of Child Psychology and Psychiatry, 43*(5), 669–678.
Wass, S. V., Jones, E. J. H., Gliga, T., Smith, T. J., Charman, T., & Johnson, M. H. (2015). Shorter spontaneous fixation durations in infants with later emerging autism. *Scientific Reports, 5,* 8284.
Wilkinson, K., Carlin, M., & Thistle, J. (2008). The role of color cues in facilitating accurate and rapid location of aided symbols by children with and without Down syndrome. *American Journal of Speech-Language Pathology, 17*(2), 179–193.
Wilkinson, K., & Jagaroo, V. (2004). Contributions of principles of visual cognitive science to AAC system display design. *Augmentative and Alternative Communication, 20*(3), 123–136.
Wilkinson, K., & Light, J. (2014). Preliminary study of gaze toward humans in photographs by individuals with autism, Down syndrome, or other intellectual disabilities: Implications for design of visual scene displays. *Augmentative and Alternative Communication, 30*(2), 130–146.
Wilkinson, K., & Mitchell, T. (2014). Eye tracking research to answer questions about augmentative and alternative communication assessment and intervention. *Augmentative and Alternative Communication, 30*(2), 106–119.
Wilkinson, K. M., & Light, J. (2011). Preliminary investigation of visual attention to human figures in photographs: Potential considerations for the design of aided AAC visual scene displays. *Journal of Speech, Language, and Hearing Research, 54*(6), 1644–1657.
Wilkinson, K. M., Light, J., & Drager, K. (2012). Considerations for the composition of visual scene displays: Potential contributions of information from visual and cognitive sciences (forum note). *Augmentative and Alternative Communication, 28*(3), 137–147.
Wilkinson, K. M., & McIlvane, W. J. (2013). Perceptual factors influence visual search for meaningful symbols in intellectual disabilities and Down syndrome or autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities, 118*(5), 353–364.
Wilkinson, K. M., & Snell, J. (2011). Facilitating children's ability to distinguish symbols for emotions: The effects of background color cues and spatial arrangement of symbols on accuracy and speed of search. *American Journal of Speech-Language Pathology, 20*(4), 288–301.
Wilson, C. E., Brock, J., & Palermo, R. (2010). Attention to social stimuli and facial identity recognition skills in autism spectrum disorder. *Journal of Intellectual Disability Research, 54*(12), 1104–1115.
Yarbus, A. L. (1967). *Eye movements and vision* (B. Haigh, Trans.). New York, NY: Plenum Press. (Original work published 1965)
Yu, C., & Smith, L. B. (2016). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science, 41*(S1), 5–31.