A child’s world is filled with many sights and sounds. In order to be successful learners, children must flexibly orient their attention to the aspects of the environment that are most relevant at a particular time. As conceptualized by Posner and Cohen (1984), visual orienting comprises three related skills: disengaging from the current focus of attention, shifting to a new stimulus, and engaging with the new stimulus. The first step in this sequence—disengaging—refers to the ability to look away from one thing in order to fixate something new. Although typically developing infants can readily disengage their attention by 3–4 months of age (Johnson et al., 1991), a growing number of studies have shown that individuals with autism spectrum disorder (ASD) show impairments in visual disengagement that emerge in infancy and last into adulthood (Elsabbagh et al., 2009, 2013; Landry and Bryson, 2004; Sabatos-DeVito et al., 2016; Sacrey et al., 2014; Zwaigenbaum et al., 2005, but see Fischer et al., 2014, 2015).

Scientists have hypothesized that impaired visual disengagement has cascading negative effects on development in children with ASD (Elsabbagh et al., 2013; Leekam and Moore, 2001; Rothbart et al., 1994), but we know little about which skills are disrupted and how this disruption takes place. This study extends our understanding by examining how individual differences in visual disengagement among children with ASD relate to a fundamental skill that has not received a great deal of consideration in this area of research—language processing. Specifically, this study examines the relationship between visual disengagement and spoken word recognition, a skill that provides a critical foundation for language development (Marchman and Fernald, 2008; Marchman et al., 2015; Weisleder and Fernald, 2013). Children with ASD vary widely in the speed and accuracy with which they process spoken words (Venker et al., 2013), and as a group they are at risk for impairments in spoken word recognition (Bavin et al., 2014). This study tests whether these
difficulties may be explained, in part, by individual differences in visual disengagement.

Although specific findings differ based on participant characteristics, task design, and stimulus salience, there is considerable evidence that individuals with ASD have more difficulty disengaging than individuals without ASD (Sacrey et al., 2014). Impaired disengagement is one of the earliest observable differences in infants who go on to receive a diagnosis of ASD, with impairments evident as early as 9–14 months of age (Elsabbagh et al., 2009, 2013; Sacrey et al., 2013; Zwaigenbaum et al., 2005). Using a computerized visual orienting task, Landry and Bryson (2004) demonstrated that these impairments are also apparent in older children with ASD (mean age 5 years), compared to children with Down syndrome and typically developing children matched on verbal and nonverbal mental age. To measure disengagement in the study by Landry and Bryson, children’s attention was first attracted to a dynamic stimulus in a central location. Once children’s attention was engaged, a new stimulus was presented on the side while the central stimulus continued; thus, switching attention to the second stimulus required disengaging from the first. Children with ASD showed striking deficits in both the speed and the likelihood of disengaging. There was almost no overlap between the mean latencies of children with ASD and children in the other groups. Furthermore, 80% of children with ASD failed to disengage on at least one trial, and children with ASD failed to disengage in 18% of trials overall.

Visual orienting is one important way in which infants and young children explore their world, and deficits in disengagement may disrupt typical developmental pathways. Elsabbagh et al. (2009) have suggested that difficulties with visual disengagement put children with ASD at risk for, “…‘locking’ onto certain irrelevant aspects of the … input” (p. 640), which in turn affects their ability to learn contingent relationships (also see Klinger et al., 2007; Renner et al., 2006). Of the many contingent relationships present in the natural world, one of the most important for young language learners is the association between spoken words (e.g. labels) and their referents (e.g. objects). As outlined in the developmental-dynamic approach to word learning (Kucker et al., 2015), one factor that helps children to build strong, correct label-object associations is effective spoken word recognition.

In the most basic sense, spoken word recognition involves orienting attention to a named object—in other words, looking at it. Experimentally, spoken word recognition is often measured using a method called looking-while-listening (Fernald et al., 2008), which presents two images on a screen with speech describing one of the images (e.g. Where’s the ball?). Children’s eye movements to the named image indicate the speed and accuracy with which they have processed the spoken noun. Although spoken word recognition is sometimes conceptualized as reflecting the endpoint of having learned a label-object association, there is evidence that effective word recognition remains key to the learning process because “… [associative learning] mechanisms build stronger links between words and objects whenever they exist together …” (Kucker et al., 2015: 3). Following this logic, looking quickly and accurately at named objects would help strengthen correct word-object links and prune incorrect ones. However, focusing attention on unnamed objects (e.g. as a result of visual orienting deficits) would build incorrect label-object associations that compete with existing associations, thus detrimentally affecting language development.

What factors determine whether a child shows good or poor spoken word recognition? First and foremost, good spoken word recognition requires language knowledge—namely, knowledge of the spoken word’s meaning (i.e. which image it describes). However, good spoken word recognition also requires that children quickly and accurately direct their gaze to the named object, which relies on non-linguistic orienting skills, including disengagement. It is therefore possible that poor visual disengagement in children with ASD disrupts the speed and accuracy with which they recognize known words, which disrupts the label-object statistics they acquire and leads to cascading negative effects on their language development. Although no studies have yet investigated the association between disengagement and spoken word recognition in children with ASD, such an approach is a critical starting point in identifying specific skills that may be disrupted by poor visual disengagement.

The majority of the published studies investigating visual disengagement in ASD have analyzed group differences, asking whether children with ASD have more difficulty disengaging than children without ASD. Building on this work, we adopted an individual differences approach by examining associations between disengagement (measured by a non-linguistic orienting task) and spoken word recognition (measured by a looking-while-listening task) in children with ASD (aged 4–7 years, n = 18). This approach is advantageous because it considers the range of proficiency in disengagement exhibited by children with ASD, with the ultimate goal of identifying developmental mechanisms. Our study was designed to address three research questions: (1) Is visual disengagement associated with the speed and accuracy of spoken word recognition in children with ASD? (2) Does visual disengagement explain unique variance in spoken word recognition after accounting for vocabulary size, a known correlate of spoken word recognition? (3) Do associations between visual disengagement and vocabulary remain significant after controlling for other developmental factors, such as age, IQ, and autism severity? We predicted that visual disengagement and vocabulary size would both
explain unique variance in spoken word recognition, even after accounting for other developmental factors.

**Method**

**Participants**

A total of 28 children with ASD diagnoses were recruited from research registries and from the community. ASD diagnoses were confirmed by the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012). Exclusionary criteria included uncorrected vision or hearing impairments, known chromosomal abnormalities, and cerebral palsy. The study took place in a child-friendly testing suite at a university research center. Parents provided written informed consent. Procedures were approved by the institutional review board (IRB). One child was excluded because he did not meet criteria for Autism or Autism Spectrum classification on the ADOS-2. Two children did not complete the ADOS-2 due to behavioral challenges; however, they were retained because they had previous ASD diagnoses and demonstrated behaviors consistent with ASD during the evaluation.

Approximately 90% of the sample contributed at least five disengage trials in the visual orienting task and at least five accuracy trials in the spoken word recognition task. To maximize the number of trials contributed by each child and increase the likelihood of obtaining a valid measure of individual children’s performance, children who contributed fewer than five disengage trials or fewer than five spoken word recognition accuracy trials were excluded (n = 7). Two additional children were excluded due to intellectual disability (Brief IQ < 70). The final sample (n = 18) was 94% male, 94% non-Hispanic, and 83% White (see Table 1). A total of 15 children met classification for Autism on the ADOS-2, and 3 children met Autism Spectrum classification.

**Standardized assessments**

The Peabody Picture Vocabulary Test, Fourth Edition (Dunn and Dunn, 2006) assessed receptive vocabulary. Growth scale values (GSVs) were used in the analyses because they measured children’s raw receptive vocabulary skills on an equal-interval scale. The Leiter International Performance Scale–Revised (Leiter; Roid and Miller, 2002) assessed nonverbal cognition. Four subtests were administered: Figure Ground, Form Completion, Sequential Order, and Repeated Patterns. Compilation of the subtests yielded a Brief IQ. The ADOS-2 confirmed ASD diagnosis and measured autism severity (ADOS-2 comparison score).

**Visual orienting task**

Children participated in a visual orienting task based on the task used by Landry and Bryson (2004). Children sat 50 in in front of a 55-in center screen in a soundproof booth. Side stimuli were presented on the periphery; children had to turn their heads approximately 65° to fully fixate the 19-in monitors on the left and right walls. Stimuli were colorful, dynamic shape patterns. To control for salience, three pairs of video clips with identical movement patterns in different colors were yoked within trials. The task included 10 shift trials and 10 disengage trials, with attention-getters interspersed every two to three trials. Children took longer to look away from the center screen (p = 0.013) in disengage trials than in shift trials, demonstrating consistency with previous studies using similar tasks. In addition, children were more likely to fail to look away from the center screen in the disengage trials than the shift trials (p = 0.005). Shift trials are not discussed further because they were not the focus of this study.

Disengagement trials presented temporally overlapping central and side stimuli. Trials proceeded as follows: the center stimulus was presented, the child had 12 s to accumulate 1 s of looking time to the center; after the child looked at the center for 1 s, the side stimulus appeared while the center video continued to play; the next trial was presented after the child accumulated 1 s of looking to the side stimulus or after the child fixated the center for 8 s (see Figure 1). Two versions were created with different trial orders; data were collapsed because performance did not significantly differ between versions (latency: p = 0.42, timeout trials: p = 0.57). Stimulus presentation was controlled by the examiner, who could see the child’s face on a screen outside the booth. The examiner pressed a button when the child was looking at the screen and released it when the child looked away. The full task lasted approximately 3.5 min.

Gaze was coded offline from video by trained coders. Trials were considered invalid and thus excluded (1) if the child was not looking at the center screen when the side

<table>
<thead>
<tr>
<th>Table 1. Participant characteristics (n = 18).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (months)</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>74 (16)</td>
</tr>
</tbody>
</table>


Brief IQ was measured by the Leiter International Performance Scale–Revised. Autism severity was measured by comparison scores on the Autism Diagnostic Observation Schedule, Second Edition.
Results

Both visual disengagement and spoken word recognition varied considerably across children. In the visual orienting task, the mean latency to disengage was $778 \pm 495$ ms, and children failed
to disengage in 11% of trials (SD=14%, range=0%–50%). The mean spoken word recognition accuracy was 80% (SD=11%, range=62%–98%), and the mean spoken word recognition latency was 554 ms (SD=192 ms, range=342 ms–989 ms).

The first research question asked whether visual disengagement was associated with spoken word recognition. We conducted a series of regression analyses with disengagement (latency or percentage of timeout trials) as the independent variable and spoken word recognition as the dependent variable. Based on our predictions, we used one-tailed p values. Disengagement latency was not a significant predictor of spoken word recognition accuracy ($b=0.02$, standard error ($SE$)=0.05, $p=0.328$) or spoken word recognition latency ($b=0.10$, $SE=0.15$, $p=0.265$), and disengagement latency was not analyzed further. The likelihood of disengaging (percentage of timeout trials) was a significant predictor of spoken word recognition, accounting for 52% of the variance in spoken word recognition accuracy ($b=-0.69$, $SE=0.17$, $p<0.001$) and 55% of the variance in spoken word recognition latency ($b=1.98$, $SE=0.48$, $p<0.001$). Figure 3 provides a visual illustration of these relationships. For simplicity, the percentage of timeout trials is referred to as “visual disengagement” in the remainder of section “Results.”

The preceding analyses were conducted again excluding one child whose percentage of timeout trials (50%) was over 2 SD above the mean, indicating a potential outlier. The percentage of timeout trials remained a significant predictor of spoken word recognition accuracy ($b=-0.85$, $SE=0.22$, $p<0.001$) and spoken word recognition latency ($b=1.88$, $SE=0.66$, $p=0.007$). The analyses for spoken word recognition latency were also conducted for the subset of $n=13$ children who contributed at least two spoken word recognition latency trials. Consistent with previous results, disengagement latency was not significantly correlated with spoken word recognition latency ($b=0.25$, $SE=0.21$, $p=0.125$), and the percentage of timeout trials was significantly correlated with spoken word recognition latency ($b=2.17$, $SE=0.76$, $p<0.001$).

The second research question asked whether visual disengagement explained unique variance in spoken word recognition after accounting for children’s vocabulary knowledge. We first conducted regression analyses

![Figure 2. Example spoken word recognition trial. In each trial, one image was the target (e.g. Where’s the ball?) and the other was the distracter. Areas of interest were defined by the edges of the gray squares containing the images, plus 10 pixels.](image)

![Figure 3. The relationship between visual disengagement and word recognition. Word recognition accuracy (left panel) was the proportion of time looking at the target image. Word recognition latency (right panel) was the time in milliseconds to look away from the distracter image. Visual disengagement was the percentage of trials in which children failed to disengage.](image)
to test whether vocabulary size was associated with spoken word recognition. As expected, vocabulary size was a significant predictor of spoken word recognition, accounting for 39% of the variance in spoken word recognition accuracy \((b=0.002, SE=0.001, p=0.003)\) and for 31% of the variance in spoken word recognition latency \((b=-0.01, SE=0.002, p=0.013)\). Next, we assessed the unique contributions of vocabulary size and visual disengagement to spoken word recognition performance by entering them as simultaneous predictors in a regression model. The results of these analyses should be interpreted cautiously given the relatively small sample size; however, previous studies have successfully used this type of approach with similar sample sizes (Bedford et al., 2014; Kaldy et al., 2011).

Both vocabulary size \((b=0.001, SE=0.001, p=0.027)\) and visual disengagement \((b=-0.53, SE=0.17, p=0.004)\) predicted unique variance in spoken word recognition accuracy, in total accounting for 63% of the variance. However, only visual disengagement was a significant unique predictor of spoken word recognition latency \((b=1.68, SE=0.60, p=0.008)\). Vocabulary size was no longer uniquely significant after accounting for disengagement \((b=-0.002, SE=0.002, p=0.201)\). In combination, vocabulary size and visual disengagement explained 57% of the variance in spoken word recognition latency. To summarize, both vocabulary size and visual disengagement accounted for non-overlapping variance in spoken word recognition accuracy, but spoken word recognition latency appeared to be more closely linked with disengagement than with vocabulary.

The third research question asked whether the associations between spoken word recognition and visual disengagement and vocabulary size remained significant after controlling for other developmental factors. Spoken word recognition, visual disengagement, and vocabulary size were not consistently correlated with IQ or autism severity (see Table 2). However, all variables of interest were significantly correlated with age. A final series of regression analyses were conducted to more precisely determine the role of age. Because of the limited sample size, it was not possible to enter all variables of interest into the same regression model. Instead, two separate sets of analyses were conducted: one with visual disengagement and age as predictors and the other with vocabulary size and age as predictors.

When visual disengagement and age were entered as simultaneous predictors of spoken word recognition accuracy \((R^2=0.57)\), only disengagement explained unique variance \((b=-0.53, SE=0.20, p=0.009)\). Similarly, when visual disengagement and age were entered as simultaneous predictors of spoken word recognition latency \((R^2=0.55)\), only disengagement explained unique variance \((b=1.77, SE=0.66, p=0.009)\). Age was not uniquely predictive in either model \((ps > 0.10)\). When vocabulary size and age were entered as simultaneous predictors of spoken word recognition accuracy, the model explained 47% of the variance in spoken word recognition accuracy. Neither age \((b=0.002, SE=0.002, p=0.08)\) nor vocabulary size \((b=0.001, SE=0.001, p=0.06)\) explained significant unique variance, although these effects were marginal. Similarly, neither age \((b=-0.006, SE=0.006, p=0.14)\) nor vocabulary size \((b=-0.004, SE=0.003, p=0.12)\) was a significant unique predictor of spoken word recognition latency; the model accounted for 37% of the variance in total. These results suggest that vocabulary size and age in large part accounted for overlapping variance in spoken word recognition, which is not surprising given that growth scale values do not account for age. In contrast, visual disengagement remained a significant predictor of spoken word recognition even after accounting for age.

### Discussion

Scientists have long suspected that deficits in visual disengagement may negatively affect development in children with ASD (Elsabbagh et al., 2013; Zwaigenbaum et al., 2005). This study advances our understanding of this issue by providing, to our knowledge, the first evidence that visual disengagement in children with ASD is associated with their proficiency in spoken word recognition. Consistent with our predictions, children with ASD with poorer visual disengagement—quantified as the percentage of trials in which children failed to disengage—demonstrated slower and less accurate spoken word recognition. Furthermore, disengagement explained significant and unique variance in spoken word recognition over and above vocabulary size, a well-known correlate of spoken word recognition (Fernald et al., 2006; Marchman and Fernald, 2008). Contrary to predictions, latency to disengage was not significantly associated with spoken word recognition, suggesting that it was the likelihood of disengaging during the full trial that mattered, not more subtle differences in the timing of disengagement.

<table>
<thead>
<tr>
<th>Table 2. Correlations.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Word recognition accuracy</td>
</tr>
<tr>
<td>Word recognition latency</td>
</tr>
<tr>
<td>Visual disengagement</td>
</tr>
<tr>
<td>Vocabulary knowledge</td>
</tr>
</tbody>
</table>

Spoken word recognition accuracy was the proportion of time looking at the target image. Spoken word recognition latency was the time (in ms) to look away from the distractor image. Visual disengagement was the percentage of trials in which the child failed to disengage. Brief IQ was measured by the Leiter International Performance Scale–Revised. Autism severity was measured by comparison scores on the ADOS-2. *p < 0.05; **p < 0.01.
Why was poorer visual disengagement associated with poorer spoken word recognition? Although it is not possible to definitively answer this question on the basis of correlational data, the most parsimonious explanation may be that the two tasks were correlated because they both relied, to some extent, on children’s ability to disengage attention. Disengagement trials presented overlapping stimuli, which required children to disengage from the center stimulus before fixating the peripheral stimulus. Spoken word recognition trials presented two images simultaneously, which required children to disengage from one image before fixating the other—a skill that was particularly important when children happened to be looking at the incorrect image when they heard the target noun (i.e. distracter-initial trials). In contrast to the non-linguistic visual orienting task, however, the spoken word recognition task presented language input to guide children’s attention—input that did not appear powerful enough to override children’s domain-general limitations in visual disengagement. It is also important to recall that children were reported to know all the words in the spoken word recognition task, meaning that a lack of disengagement could not be explained by a failure to understand which image the label described.

If poor visual disengagement does, in fact, disrupt the speed and accuracy of spoken word recognition in children with ASD, such disruptions would likely lead to cascading negative effects on language development. Locking onto irrelevant aspects of the environment would be detrimental because the language input a child hears would not necessarily relate to what he sees at any given moment. This lack of alignment could produce weak lexical representations even for words a child knows. Such detrimental effects could be even more disruptive for words that are less familiar or that a child is just beginning to learn. Failing to look at the right thing, or looking too late, could affect what children learn about language and how quickly they learn it—particularly when they are faced with ambiguous learning contexts that require integration of multiple object-label co-occurrences over time (Smith and Yu, 2008; Yu and Smith, 2011). Future work is needed to determine the effects of impaired disengagement on word learning.

Although additional research is needed to directly test this hypothesis, the current results are consistent with a developmental model in which differences in non-linguistic visual orienting skills negatively affect children’s ability to quickly and accurately recognize spoken words, which subsequently affects their language development. Consistent with this idea, Keehn et al. (2013) proposed a model in which impaired disengagement disrupts language development by reducing or disrupting attention shifting. The fact that deficits in disengagement appear and worsen during the period of development in which children typically learn a great deal of language—the second year of life—makes this an even more provocative area for further investigation. Longitudinal studies of young children with ASD are needed to determine how relationships among visual disengagement, spoken word recognition, and language change over the course of early development.

These findings suggest that in order to fully understand why some children with ASD are at risk for impairments in spoken word recognition, we must consider their non-linguistic attention—specifically, visual disengagement. Knowing how visual disengagement relates to language may help us understand why some children with ASD develop age-appropriate language skills, but others experience lasting difficulties. If future studies confirm that poor visual disengagement interferes with language processing and learning in children with ASD, it will be necessary to broaden theories of language development in children with ASD to incorporate not only the influence of social attention on language development but also the influence of domain-general aspects of non-social attention (see Bedford et al., 2014).

It would be ideal to treat impairments in visual disengagement as they emerge, potentially preventing a negative developmental cascade. Although it is rare for interventions to directly target attentional behaviors such as disengagement (Patten and Watson, 2011), one promising strategy is using cues to orient children’s attention (Koegel et al., 2009; Walton and Ingersoll, 2013). There is also evidence that short-term attentional training can reduce disengagement latencies in typically developing infants (Wass et al., 2011), suggesting avenues for research in children with ASD. In time, this line of work may inform the development of novel intervention strategies, as well as a clearer understanding of how these strategies work.

Acknowledgments
I thank Susan Ellis Weismer, Jenny Saffran, Anthony Goodwin, Erica Wojcik, and Viridiana Benitez for providing feedback on previous drafts of this manuscript. Thanks to Anna Dorrance, Taryn Stricker, and Kristina Fassbender for assistance with participant visits and coding. Thanks to the members of the Language Processes Lab and the Infant Learning Lab for feedback, especially Sara Kover and Casey Lew-Williams. Sincere thanks to the participants and their families.

Funding
This work was supported by F31DC012451 (Venker, PI), R01DC011750 (Ellis Weismer, PI, Kaushanskaya, co-PI), R37HD037466 (Saffran, PI), P30HD003352 (Mailick, PI), the Friends of the Waisman Center, and the Wisconsin Speech-Language Pathology and Audiology Association.

References

{References}


