The role of intention in reading referential gaze:

Implications for learning in typical development and in Autism Spectrum Disorder

Janet Y. Bang
School of Communication Sciences and Disorders
Faculty of Medicine
McGill University, Montreal
February, 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Communication Sciences and Disorders

To all new students who are struggling with seeing the light at the end of the tunnel, don’t worry, you’ll get there!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TABLES IN APPENDICES

LIST OF FIGURES IN APPENDICES

ABSTRACT

A ubiquitous cue in our interactions with children is *referential gaze* (when a person looks at an object). For example, children can learn a new word by listening to someone say the name of an object and following his/her referential gaze, a *word learning context*. Children can also predict what object someone will use next by following his/her referential gaze, an *action prediction context*. Referential gaze is important because we can learn about our environment from this subtle cue. It is often proposed that children treat referential gaze as a cue that conveys another's communicative intent, and thus they are able to learn new words or predict another's actions. However, another possibility is that children simply treat referential gaze as a cue to attend to something. It is critical to understand how children treat referential gaze, particularly if there are children who may use referential gaze less effectively, such as children with autism spectrum disorder (ASD).

I tested the intentionality of gaze in two ways. First, by comparing children's attention to and learning from a directional cue that is not likely to be intentional, an arrow cue, with the potentially intentional referential gaze cue. Cues were matched as closely as possible on size, motion, and physical features such that differences in responding to cues could indicate that gaze is treated as intentional. Second, I compared performance between cues in two different populations: typically-developing children (TD) and children with ASD. Because of their social impairments, children with ASD are proposed to have difficulties with treating referential gaze as intentional. However, in both populations, referential gaze is often assumed to be intentional in contexts of word learning and action prediction, with no studies comparing performance between a referential gaze cue and a closely matched control cue.

Children were matched on nonverbal IQ, chronological age, ratio of girls to boys, and parental education (ASD n = 24, TD n = 24). Children watched videos of either an arrow or a gaze cue indicating a target object. In the word learning context, children were taught the name of the target object and in the action prediction context, children decided whether the target object (a block) should complete a block tower. Performance measures included visual attention and *in-depth learning measures* that assessed what children learned beyond the simple associations taught in the video. In-depth learning was assessed both immediately after teaching and one week later. Results revealed no group differences, but cue condition differences in both contexts. In the word learning context, both children with ASD and TD children looked more at the area of gaze, were faster to locate the target object, and recalled more semantic features about the target object when the cue was gaze versus an arrow. In the action prediction context, only TD children looked more at the area of gaze, though neither group of children showed evidence of learning with either arrow or gaze cues.

This study is the first to provide evidence that suggests children with ASD and TD children may use referential gaze as an intentional cue, rather than simply a directional cue in a word learning context. Moreover, an intentional reading of gaze by children in both groups may have positive implications for what they recall about an object. In contrast, in the action prediction context, neither group of children demonstrated learning, which may have been due to difficulties inherent to the task. Future studies should continue to investigate the relationship between intention understanding and its implications for children's learning.

RÉSUMÉ

Un signal omniprésent dans nos interactions avec les enfants est *le regard référentiel* (quand une personne regarde un objet). Par exemple, les enfants peuvent apprendre un nouveau mot en écoutant quelqu'un dire le nom d'un objet et en suivant son regard référentiel : *un contexte d'apprentissage de mots*. Les enfants peuvent également prédire quel objet quelqu'un utilisera ensuite en suivant son regard référentiel : *un contexte de prédiction d'action*. Le regard référentiel est important parce que nous pouvons apprendre de notre environnement à partir de ce signal subtil. Il est souvent proposé que les enfants considèrent le regard référentiel comme un signal qui traduit l'intention communicative d'un autre, et qu'ainsi, ils sont capables d'apprendre de nouveaux mots ou de prédire les actions d'autrui. Cependant, une autre possibilité est que les enfants traitent simplement le regard référentiel comme un signal pour prêter attention à quelque chose. Il est essentiel de comprendre comment les enfants traitent le regard référentiel, à savoir si certains enfants utilisent le regard référentiel de manière moins efficace, comme les enfants avec un trouble du spectre autistique (TSA).

J'ai examiné l'intentionnalité du regard de deux façons. Premièrement, j'ai comparé les capacités d'attention et d'apprentissage des enfants face à un signal directionnel non susceptible d'être intentionnel (un signal de flèche) à leurs capacités d'attention et d'apprentissage face à un signal potentiellement intentionnel (le regard référentiel). Les signaux ont été créés les plus équivalents possibles, pour s'égaler en termes de taille, de mouvement et de caractéristiques physiques, de sorte que des réponses différentes aux signaux ne puissent être attribuées qu'à l'intentionnalité du regard. Deuxièmement, j'ai comparé la performance d'enfants de deux populations différentes : des enfants à développement typique (DT) et des enfants avec un TSA. Il est estimé qu'en raison de leurs déficiences sociales, les enfants avec un TSA ont des difficultés à traiter le regard référentiel comme intentionnel. Cependant, dans les deux populations, le regard référentiel est souvent considéré intentionnel dans les contextes d'apprentissage de mots et de prédiction d'action, sans aucune étude à ce jour ayant comparé la performance des enfants devant un regard référentiel à celle devant un signal de contrôle apparié.

Les enfants ont été comparés sur le QI non verbal, l'âge chronologique, le rapport filles-garçons et l'éducation des parents (DT n = 24, TSA n =24). Les enfants ont regardé des vidéos présentant un objet cible avec un signal de flèche ou un regard. Dans le contexte d'apprentissage de mots, les enfants devaient apprendre le nom de l'objet cible, tandis que dans le contexte de prédiction d'action, ils devaient décider si l'objet cible (un bloc) devrait compléter une tour de blocs. Des mesures d'attention visuelle et des *mesures d'apprentissage approfondi* qui évaluent ce que les enfants ont appris au-delà des simples associations enseignées dans la vidéo, ont servi pour mesurer le rendement des enfants. L'apprentissage approfondi a été évalué immédiatement après l'enseignement et une semaine plus tard. Les résultats n'ont révélé aucune différence entre les groupes, tandis que des différences entre les signaux ont été trouvées dans les deux contextes. Dans le contexte d'apprentissage de mots, les enfants avec un TSA et à DT ont davantage regardé la zone du regard, étaient plus rapides pour localiser l'objet cible et se rappelaient davantage des caractéristiques sémantiques de l'objet cible lorsque le signal était le regard par rapport à la flèche. Dans le contexte de prédiction d'action, seuls les enfants à DT ont plus regardé la zone du regard, bien que ni l'un ni l'autre des groupes n'ait montré d'apprentissage avec la flèche ni le regard.

Cette étude est la première à fournir des preuves suggérant que les enfants avec un TSA et à DT peuvent utiliser le regard référentiel comme signal intentionnel, plutôt qu'un simple signal directionnel dans un contexte d'apprentissage de mots. En outre, une interprétation intentionnelle du regard par les enfants des deux groupes peut être bénéfique pour intégrer davantage d’information à propos d’un objet. En revanche, dans le contexte de prédiction d'action, aucun des groupes n'a démontré d'apprentissage, ce qui peut être attribuable à des difficultés inhérentes de la tâche. Les études futures devraient poursuivre l’examen de la relation entre l'intention et les implications pour l'apprentissage des enfants.

ACKNOWLEDGEMENTS

This dissertation is my attempt at bringing together the knowledge I have gained over all of the friendly discussions (some, admittedly, over the necessary beer), lectures, and seminars over the past seven years. None of this would have been possible without numerous people that I have had the pleasure of working with here at McGill.

To my supervisor, Dr. Aparna Nadig, you have challenged me to think critically about research and your encouragement has kept me moving forward every time I felt discouraged or disheartened (with the occasional exchange of chocolates to get us through midday and late afternoon meetings). I cannot thank you enough for giving me the opportunities to learn about all facets of research, and I have especially appreciated your thoughtful comments and critiques dedicated to helping me be a stronger writer. I feel that I am very fortunate to have found someone who I could count on to support me during this process.

I would like to thank my dissertation committee members, Dr. Linda Polka and Dr. Kristine Onishi. Your insightful feedback on this work from the day of the oral presentation to the written dissertation has helped me think from new perspectives and made this work even stronger. I would also like to thank Dr. Onishi for all of your helpful comments on the videos and initial design of the study, as well as the detailed feedback on the written dissertation. Thank you to Dr. Yuriko Oshima-Takane and Dr. Athena Vouloumanos, who, along with Dr. Onishi, provided the testing rooms and the eye tracker that I used in this project. I also greatly appreciated the feedback provided by Dr. Armando Bertone during the initial design of the study.

A big thank you to the organizations that helped us recruit including the Montreal Children's Hospital, Chantal Mongeau au Centre intégré universitaire de santé et de services sociaux du Centre-Sud-de-l'Île-de-Montréal (CIUSSS), Krista Leitham and the Autism Speaks

Canada Walk in Montreal, Summit School, Giant Steps, and Autisme Montréal. A very grateful thank you to my fellow lab mates who provided encouragement when I thought that I would never start testing, and when I did start testing, continued to provide encouragement while discussing our musings on research: Hanady Bani Hani, Ana Maria Gonzalez Barrero, Jesse Burns, Céliane Trudel, and Melanie Eberhardt. To my dedicated RAs, you all have really seen me through the heart of this project, working with the families. I cannot thank you enough for the time you committed through all the early morning sessions, walking through snowstorms to home visits, and staying at the lab on evenings and weekends for data entry and testing. You all helped keep me sane: Corentin Montiel, Milva Venditti, Verona Soliman, Hida Caliskan, Lisa Bissett, Nicole Khammar, Renuka Giles, Nowrin Hoque. Thank you to Marta Gunin for always being such a positive supporter of my French and translating the abstract. I am also indebted to Melissa Ferland for putting me in touch with Isabelle Leduc-Cummings, who was the most patient actress while making the video and audio stimuli. Most importantly, I am eternally grateful to all of the families that took the time out of their busy schedules to participate in this study in rain, snow or shine. This work would not have been possible without your support.

I would also like to thank members of the various reading groups, the SCSD writing group, the SCSD statistics group (especially to Stefanie Nickels in our efforts to understand the nitty gritty of statistics), and the ART Journal Club (especially the lifers: Jackie Guy, Ana Tryfon, and Megha Sharda). The discussions from these groups have given me a seemingly never-ending list of new questions, but still managed to resolve some questions along the way.

To all of the friends I've made while at McGill and to those who have moved onto new chapters in their lives, thank you for the good times that helped make Montreal a second home. To Kent Mori and the Maderight team, thanks for sharing your office space during this writing process and interjecting, more often than not, ridiculous commentaries on life that gave me a healthy diversion from staring at my computer all day. I would also like to thank Sara Palencia for lending her wonderfully opinionated and supportive ear to my ramblings throughout this process. To my friends outside of research, I always know you will support me no matter what I do, and for that I am truly thankful. Lastly, thank you to my family and my new extended family for always keeping me focused in life!

PREFACE AND CONTRIBUTION OF AUTHORS

This dissertation is the first to examine whether children with autism spectrum disorders (ASD) and typically-developing (TD) children use referential gaze as a directional cue, or potentially, an intentional cue, in contexts of learning. In contexts of word learning and action prediction, learning was compared between an arrow (a control for the direction of attention) and a referential gaze cue; differences between cues were interpreted as children treating referential gaze as an intentional cue. No other studies to date have compared learning between two different cues in the same children within two populations of those with ASD or TD. Moreover, this study is the first to examine children's immediate and long term in-depth learning (i.e., semantic knowledge and the ability to generalize learning) when learning from an arrow or a referential gaze cue. In the context of word learning, the findings from this study suggest that in contrast to prior research in ASD, that both children with ASD and TD children may be treating gaze as an intentional cue to help them learn new words, and this intention understanding may have benefits for their recollection of what an object looks like and how it can be used. In the context of action prediction, the findings from this study suggest learning with either cue may be difficult.

This dissertation is written in a standard format and consists of two behavioral studies, one in the context of word learning and the other in the context of action prediction. I developed the hypotheses and design of the study with input from my supervisor, Dr. Aparna Nadig. An oral presentation of the proposal was presented to my dissertation committee members, Dr. Kristine Onishi and Dr. Linda Polka, and all committee members provided additional feedback on the proposal. Dr. Kristine Onishi provided valuable additional feedback on research methods over the course of the study design. The experiments were conducted with the support of seven

undergraduate research assistants, Hida Caliskan, Corentin Montiel, Verona Soliman, Milva Venditti, Lisa Bissett, Nowrin Hoque, Renuka Giles, and Nicole Khammar, who assisted with object creation, scheduling, recruitment, data entry, data organization, and coding. I made all of the videos, and the eye tracking data could not have been collected without the eye tracker and testing space of Dr. Kristine Onishi, Dr. Yuriko Oshima-Takane and Dr. Athena Vouloumanos. I completed all data analyses and writing of the dissertation. My supervisor provided feedback from the point of my initial drafts, and I received valuable feedback from dissertation committee members on later drafts of the dissertation. Some of the findings from this dissertation were presented at the following conferences: Child Language Symposium, Coventry, United Kingdom 2015; International Meeting for Autism Research, Baltimore, Maryland, United States 2016; workshop on “The role of pragmatic factors in child language processing”, Berlin, Germany 2016.

I was funded for this work by the Canadian Institutes of Health Research through the Autism Research Training Program as a Strategic Training Fellow with supplemental funding from the Sinneave Family Foundation, The Center for Research on Brain, Language and Music’s Graduate Scholar Stipend, and a Graduate Excellence Fellowship from the School of Communication Sciences and Disorders. Additional funding was provided by Dr. Nadig from the Social Science and Humanities Research Council of Canada and the MaxBell Foundation.

GLOSSARY

The following are a list of terms that are used frequently throughout my dissertation. These terms are defined in the way that they are used in this dissertation.

Action prediction task	A scenario where children are asked to guess an actor's next unknown action or goal
Arrow	A bar with a point on one side (a sideways triangle) that conveys directional information
Autism Spectrum Disorder (ASD)	A neurodevelopmental disorder defined by two major deficits: impairments in social communication and interaction, and the presence of restricted, repetitive behaviors and stereotyped interests (APA, 2013)
Context	A scenario or a situation; in this dissertation I refer to two different contexts of word learning and action prediction
Cue	An act or sign that conveys information; this dissertation uses this term to refer to cues of referential gaze and an arrow
Developmental delay	Children without ASD who have delays in their development such as in their cognition and/or language (e.g., Down Syndrome)
Fast-mapping	When a label-object association is learned through minimal exposure of the label being presented in conjunction with an object
In-depth learning measures	Measures that go beyond a simple label-object association; in this dissertation in-depth learning measures examine children's semantic knowledge and how children can generalize what they learned (i.e., generalize a label to other objects that are similar in kind or generalize which block will be used to build a tower)
(Communicative) Intention	The understanding that another person communicates information in order to share it with someone else

Lean interpretation	The assumption that lower level processes (attention) explain the pattern of data; in this dissertation I discuss if a lean or rich interpretation is plausible for how children treat referential gaze
Referential gaze	When a person looks at an object or event
Rich interpretation	The assumption that in addition to lower-level processes, understanding mental states is necessary to explain the pattern of data; in this dissertation I discuss if a lean or rich interpretation is plausible for how children treat referential gaze
Semantic knowledge	The meaning of a word; in this dissertation the meaning is defined by what an object looks like and how it can be used
Typical Development/ Typically-developing (TD)	Children who do not have ASD, or any other developmental learning or behavior disorder
Word learning task	A scenario where children are taught a new word and asked to demonstrate if they have learned this word

GENERAL INTRODUCTION

A Lean Versus Rich Interpretation of Referential Gaze

Imagine that you are building a block castle with a young child named Ben. You tell Ben that you will teach him some names you know. While Ben is looking at a piece in his hand, you notice an object on the floor called a portcullis (i.e., the sliding gate for the front of the castle) lying next to other pieces. You look up at Ben and say, "*That's a portcullis*!" then look back at the object. This looking at the object is also known as providing *referential gaze*. You then look up at Ben and ask him, "*Can you hand me the portcullis?*", and he reaches specifically for the object you looked at. This *context,* or scenario, demonstrates *word learning*, where Ben learned the name of the object by following your referential gaze.

Next, you begin to build the castle tower by stacking blocks with Ben. Ben looks at a flag to put on top of the tower, but you look up at Ben and provide referential gaze to a brown block closer to Ben than to you. You then ask Ben, "*Can you give me that one?*". Ben reaches for the brown block, rather than giving you the flag he was looking at. This context demonstrates *action prediction*, where Ben predicted which block you were going to use next by following your referential gaze. The type of learning differs between the word learning and the action prediction contexts. While in word learning, Ben learns a name for an object, in action prediction, Ben learns your preference for which block to use next. However, in both contexts, Ben's learning was achieved by using your referential gaze. In this dissertation, I investigate how your referential gaze contributed to Ben's learning in both contexts.

There are two opposing interpretations that can explain how Ben used your referential gaze to learn in these contexts. On one end, a *lean* interpretation would be that Ben used the direction of your gaze as a cue to help him learn. Thus learning can be explained by Ben aligning attentional and spatial features such as your head turn, your direction of gaze and the object (Heyes, 2014a; 2014b; Leekam, 2016). On the other end, a *rich* interpretation would be that Ben learned because he read beyond the direction of your gaze, and understood your gaze as a cue that reflects your communicative intention (hereby referred to as *intention*; Baldwin, 1993a; Baron-Cohen, 1995; Meltzoff, 2007; Tomasello, 1999). Therefore, Ben learned because he understood your intention to communicate and share information with him about that object. There is also evidence for a developmental shift, where young children may use attentional and spatial features to learn initially, but may shift to intention understanding around the age of 2 or 3 years (Doherty, 2006; Hollich et al., 2000; Moore & Corkum, 1994; Woodward, 2003). Importantly, some supporters of a rich interpretation have not only proposed that gaze is intentional, but that this intention understanding contributes important positive benefits to what we learn about the world around us (Becchio, Bertone, & Castiello, 2008; Csibra & Gergely, 2009).

While many studies have provided evidence to support a rich interpretation of referential gaze, few studies have provided strong enough evidence to eliminate the possibility that learning may have occurred without treating gaze as intentional. In other words, studies often assume that learning with referential gaze is because of intention understanding, but few studies have tested whether the same learning can occur by simply following the direction of referential gaze (Field, 2016). One way to tease apart the directional and intentional reading of gaze is to compare learning with a control cue, that only provides directional information and not intentional information, versus a referential gaze cue. A common directional cue that has been used as a control for referential gaze is an arrow (e.g., Birmingham & Kingstone, 2009). An arrow cue can be matched as much as possible on size, motion, and physical features (e.g., the color contrast of the sclera and pupil), such that any differences in responding to an arrow versus a referential gaze cue may, in part, be because individuals treat referential gaze as intentional. It is possible that an arrow cue could be considered intentional, in that we think someone intended to use this cue to convey direction, but it is not used in constant social interaction in the same way as referential gaze. Thus for the purpose of this dissertation I will consider an arrow cue as directional, and not intentional. If children show differences in responding to referential gaze versus an arrow, I interpret this difference as referential gaze being treated as an intentional cue, different from a simply directional cue of an arrow. If children similarly respond to both cues, these similarities could suggest that children treat referential gaze as directional. However, there may be other possible conclusions such as both cues are similarly salient, or interesting, or that potential differences in learning between cues could not be detected in the sample. Currently, it is unclear as to whether referential gaze contributes to learning because it is an effective directional or intentional cue, and the use of adequate controls is critical to helping us understand this question. Determining how referential gaze contributes to learning is important if some children are not able to use referential gaze effectively to help them learn.

One population for which this is particularly relevant is children with *autism spectrum disorder (ASD)*. ASD is a neurodevelopmental disorder defined in part by impairments in social communication and interaction. Some researchers have proposed that because of impairments in the social domain, children with ASD lack the ability to read gaze as an intentional cue (Baron-Cohen, Baldwin, & Crowson, 1997; Norbury et al., 2010; Preissler & Carey, 2005). Yet there are two main assumptions that are prevalent in ASD research that make it difficult to interpret findings regarding an intentional reading of referential gaze. First, researchers often assume that typically-developing children treat gaze as intentional. Then, when these studies find impairments in children with ASD relative to children without ASD (e.g., typically-developing children or children with developmental delay who do not have ASD), researchers assume that because of their known social impairments, children with ASD do not understand intention.

I address these assumptions by comparing learning from an arrow and a referential gaze cue in both groups of children with ASD and typically-developing children. The comparison of both cues in both groups of children would reveal within- and between-group differences in how children treat referential gaze, and the evidence can be interpreted without relying on the aforementioned assumptions in either group. As mentioned above, in this dissertation I am interested in how referential gaze contributes to learning in both contexts, and I investigate this question by comparing whether children with ASD and typically-developing children treat referential gaze as a directional or an intentional cue. Learning will be examined separately in the contexts of word learning (Chapter 1) and action prediction (Chapter 2). The following section provides an overview of the role of referential gaze in ASD research.

Autism Spectrum Disorder and Referential Gaze

ASD is a neurodevelopmental disorder defined by two major domains: impairments in social communication and interaction, and the presence of restricted, repetitive behaviors and stereotyped interests (American Psychiatric Association, 2013). While these domains are known to be impaired in individuals with ASD, it is well understood that on other domains there is a wider range of individual differences in children's level of functioning from impairment to similar levels as typically-developing individuals. For example, while one individual may have no functional language, another individual could have structural language abilities (as opposed to pragmatic language abilities) that are no different from typically-developing children (Kjelgaard

& Tager-Flusberg, 2001; Pickles, Anderson, & Lord, 2014; Tager-Flusberg, 2006). Other domains such as IQ have shown a similarly wide range in children with ASD (Charman et al., 2011; Tager-Flusberg & Joseph, 2003). However, despite this heterogeneity, one defining characteristic in this population is the impairment in social communication and interaction.

One key component of social communication and interaction is how we attend to and use another's eye gaze (for review see Boraston & Blakemore, 2007; Itier & Batty, 2009; Nation & Penny, 2008; also see Weigelt, Koldewyn, & Kanwisher, 2012 for a review of attention to faces in ASD research). In children with typical development, following another's spontaneous gaze direction can be seen as early as 10 to 12 months (Butterworth & Jarrett, 1991; Moore & Corkum, 1998; Scaife & Bruner, 1975). Yet in children with ASD, an impairment in following spontaneous gaze direction is one of the earliest clinical markers (Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997; Loveland & Landry, 1986). Spontaneous gaze following is important because it is an early precursor to joint attention abilities, which is the coordination of triadic attention between two people, and an object or event. The ability to engage in joint attention is considered early evidence of children understanding another's mental state, or theory of mind (e.g., Baron-Cohen, 1995). Theory of mind refers to an awareness that mental states exist in yourself and others, such as intentions, desires, and beliefs (Premack & Woodruff, 1978). One of the most prevalent psychological theories of why individuals with ASD have difficulties in social communication and interaction is that they lack a theory of mind (Baron-Cohen, Leslie, & Frith, 1985, but see also Astington & Jenkins, 1999; Happé, 1995 for limitations of this account). Moreover, the ability to detect and use another's eye gaze has been proposed as a critical step in the development of one's theory of mind (Baron-Cohen, Campbell, Karmiloff-Smith, Grant, & Walker, 1995).

Initial studies that examined referential gaze and theory of mind concluded a rich interpretation of gaze. In one study, Baron-Cohen and colleagues (1995) examined whether individuals with ASD (chronological age range 8;0 to 18;2 years, language age range 4;0 – 5;8 years) could determine the mental state of Charlie, a schematic face, based on the direction of his eyes. Control groups included individuals with *developmental delay (DD)* matched on chronological and language ages to those with ASD (DD chronological age range 5;0 – 21;5 years) and typically-developing individuals matched on language age to those with ASD (TD chronological age range 4;0 – 4;8 years). In their four sweets task, the authors presented Charlie at the center of a transparency, and overlaid a sheet with a picture of a candy on each corner. With Charlie's eyes directed at one of the four corners, the authors asked participants questions about what Charlie wanted (desire; "*Which one does Charlie want?*"), what he was going to take (goal; "*Which one is he going to take?*"), and his intention to name a candy ("*One of these is a beb. Which one is the beb?*"). The authors found that individuals with ASD, in contrast to control groups, had difficulty using Charlie's eyes to determine his mental state. Because control groups were able to determine Charlie's mental state, the authors concluded that the ASD group demonstrated an impaired understanding of mental states.

This conclusion relies on the assumption that difficulty on the task reflects an impairment in understanding mental states. It is possible, however, that individuals with ASD had difficulty because of other lower-level explanations, such as a difficulty with attention and sensory processing (Leekam, 2016), for example, the direction of Charlie's gaze. Even in typical development, Heyes (Heyes, 2014a; 2014b) advised that instances of *mentalizing*, or thinking about another's mental states, may be better explained by lower-level processes. Heyes proposed that to identify potentially true instances of mentalizing, one method would be to provide evidence of whether a control condition, that does not employ an animate being such as a person, could also elicit the supposed mentalizing behavior.

Baron-Cohen and colleagues (1995) did use a control condition in their experiment 3 by including an arrow cue that pointed to another candy. However, the condition was designed with competing cues where Charlie looked at one candy while an arrow cue pointed at another candy. The authors compared the mean number of times children followed Charlie's gaze, and the types of errors when children did not follow Charlie's gaze. Baron-Cohen et al. found that in comparison to control groups, children with ASD followed Charlie's gaze less and made more errors that followed the arrow. The authors used this data to conclude that children with ASD may be "blind to the mentalistic significance of the eyes" (p. 394); see Rombough and Iarocci (2013) for similar findings. While a preference for gaze within competing cues has been interpreted as evidence that gaze is an intentional cue (Aldaqre, Paulus, & Sodian, 2015; Field, 2016; Moore, Angelopoulos, & Bennett, 1999), one problem with competing cues is that it forces children to choose between cues, without an understanding of how a child responds to each cue presented in isolation. Children may have difficulty or be successful with both cues in isolation, but the combination of cues may be confusing. In contrast to competing cues, comparing performance with gaze in one condition and an arrow in another condition can provide information regarding whether both cues are similarly effective, or similarly difficult for children, or if there is in fact a difference between groups in how they treat a directional cue such as an arrow or a potentially intentional cue such as referential gaze.

In their experiment 2, Ames and Jarrold (2006) used the desire question of the four sweets task but included Charlie's gaze in one condition and an arrow in another condition. The authors included multiple trials that were divided into two windows of early and late trials, and children with ASD as well as two control groups (typically-developing children, children with learning difficulties who do not have an ASD). While control groups were similar in their responding to gaze and arrow cues, children with ASD showed differences between cues only in early trials. In early trials, children with ASD responded more with an arrow versus a gaze cue, whereas in late trials there was no difference between cues. However, relative to control groups, children with ASD were still impaired in early trials with the gaze cue and in late trials with both the gaze and arrow cues, suggesting that lower-level directional information provided by gaze and arrow cues was still difficult for children with ASD. These findings suggest that successful responding to Charlie's gaze may in part be due to following lower-level directional information. This interpretation could not have been observed when cues were in competition, highlighting the importance of comparing each cue in isolation when studying referential gaze.

Attention research has compared referential gaze relative to an arrow cue in two different conditions using spatial cueing paradigms (for reviews see Birmingham, Ristic, & Kingstone, 2012; Nation & Penny, 2008; Rombough, Barrie, & Iarocci, 2012). Spatial cueing paradigms examine how children orient their attention in response to a central cue. Evidence from these paradigms contributes to the larger question of how children follow or orient to gaze, but the paradigms differ from gaze following in contexts of word learning or action prediction. Spatial cueing paradigms are presented on a computer and assess repeated, rapid, automatic responses to a cue, often with over a hundred trials. Gaze and arrow cues are both common cues used in this paradigm. On each trial, a central cue appears and points to the left or the right side of the screen where a target image will subsequently appear. If participants are faster to respond to the target when it appears at the cued location versus when the target appears at the uncued location, then this difference in cueing response is considered as an involuntary, automatic response to the cue.

Both gaze and arrow cues have demonstrated this cueing response, in both individuals with ASD and typically-developing individuals (Kuhn et al., 2010; Pruett et al., 2010; Tipples, 2002; 2008), making it difficult to conclude whether gaze is treated as directional, or perhaps even intentional for both groups of children, or if there are differences that could not be detected in these studies. Yet other spatial cueing studies have found that despite some similar cueing responses, typically-developing individuals do treat a gaze cue differently from an arrow cue (Hietanen, Nummenmaa, Nyman, Parkkola, & Hämäläinen, 2006; Marotta, Lupiáñez, & Casagrande, 2012a; Ristic, Friesen, & Kingstone, 2002; Ristic, Wright, & Kingstone, 2007), and individuals with ASD do not treat arrow and gaze cues differently (Senju, Tojo, Dairoku, & Hasegawa, 2004; Vlamings, Stauder, Son, & Mottron, 2005). The lack of differences between cues in individuals with ASD relative to cue differences seen in typical development have been interpreted as typically-developing individuals reading gaze as an intentional cue, whereas individuals with ASD read gaze as a simple directional cue. A number of reasons for these mixed findings include differences in the type of stimuli used (e.g., images, schematic faces), or the high variability in response to gaze in individuals with ASD (Birmingham et al., 2012; Rombough et al., 2012). While data from these studies contribute evidence for how children follow or orient to gaze, varied findings make it difficult to draw clear conclusions. Additionally, these paradigms do not address gaze following in contexts of learning.

This dissertation addresses a critical gap in the literature by investigating whether an intentional reading of gaze can have implications for children's learning. Specifically, I focus on contexts of word learning (Chapter 1) and action prediction (Chapter 2) in 6- to 11-year-old children with ASD and typically-developing children. I examine if learning in each context differs when children learn from an arrow cue (i.e., control for the direction of attention) or a potentially intentional, referential gaze cue

Additionally, in both word learning and action prediction, I include measures that are aimed to tap into multiple aspects of learning. For example, Ben demonstrated some indication of learning the word *portcullis* when he reached for the object, but this only demonstrated one aspect of word learning, where he was able to receptively associate the object he saw with the label. Other measures could address more *in-depth learning,* which refers to other aspects of learning that demonstrate that Ben learned beyond the simple association he was taught (e.g., a simple association between the label and the object he saw). For example, in the context of word learning, in-depth learning measures could demonstrate an understanding of the meaning about the word *portcullis,* such as describing the word, or generalizing the word *portcullis* to objects that are similar in kind (e.g., different color from the first portcullis Ben saw). In this dissertation, I compare in-depth learning between an arrow cue (control for the direction of attention) relative to a referential gaze cue in both children with ASD and typically-developing children, to examine if intention understanding has implications for children's in-depth learning. Moreover, I examine the retention of learning one week later, which speaks to the implications for the consolidation of learning over time (Dumay & Gaskell, 2007; Henderson, Weighall, Brown, & Gareth Gaskell, 2012; Henderson, Powell, Gareth Gaskell, & Norbury, 2014; Norbury et al., 2010).

In summary, referential gaze is used frequently throughout our daily interactions, and it is often assumed to be a rich communicative and intentional cue (Baron-Cohen, 1995; Meltzoff, 2007). However, these claims cannot be supported until referential gaze is tested against non-intentional cues, such as an arrow. If an arrow cue is controlled for as many features of gaze as possible, differences in how children learn in response to a gaze or an arrow cue could, to a certain extent, be due to an intentional reading of gaze. Though comparisons of referential gaze and arrow cues have been seen in spatial cueing research[1], findings are mixed and more importantly, do not address how either children with ASD or typically-developing children treat referential gaze in contexts of learning. If children with ASD do not treat referential gaze as intentional, but typically-developing children do, and treating gaze as an intentional cue strengthens learning, then children with ASD may not be privy to the same benefits of learning from referential gaze that may be seen by their typically-developing peers. By investigating the role of referential gaze in contexts of learning, and extending this investigation into in-depth learning in each context, this dissertation addresses the fundamental understanding of referential gaze as a directional or intentional cue. This understanding is especially important if children with ASD do not treat gaze as intentional, and a lack of intention understanding may have negative implications on their in-depth learning.

Each of the following chapters include an Introduction, Methods, Results, and Discussion section in the respective context of learning. At the end of Chapter 2, learning across contexts is compared to examine learning from cues across different contexts. The General Discussion offers final thoughts, Limitations, and Future Directions on the directional versus intentional reading of referential gaze in word learning and action prediction contexts.

---

[1] Often the primary feature of gaze that is controlled for in these studies is direction, with the control of an arrow cue. Studies differ in how they control other features such as color, complexity, motion, and texture.

CHAPTER 1: REFERENTIAL GAZE AND WORD LEARNING

INTRODUCTION

Ben learned the word portcullis through minimal exposure to a new object being labeled, also referred to as fast-mapping (Carey & Bartlett, 1978). Specifically, Ben's fast-mapping was facilitated by following your referential gaze. Yet is unclear why your referential gaze contributed to his word learning. A lean interpretation would be that he simply followed the directional cue of your gaze. In contrast, a rich interpretation would be that he used your referential gaze as an intentional cue. Neither interpretation actually changes what Ben learned. However, it is important to know how children with ASD treat referential gaze in word learning contexts, if an inability or difficulty with treating gaze as an intentional cue negatively impacts their learning of new words.

What does the literature conclude given that Ben has an ASD? Some research has suggested that individuals with ASD cannot understand intent in the context of word learning (Baron-Cohen et al., 1997; Norbury et al., 2010; Preissler & Carey, 2005). Other research points to the possibility that children with ASD can understand intention for word learning (Aldaqre et al., 2015; Bani Hani, Gonzalez-Barrero, & Nadig, 2012). However, these studies have based these conclusions on the assumption that successful learning assumes intention understanding, and no studies to my knowledge have compared a control for referential gaze against referential gaze itself for the same participants. The current study compares how individual children with ASD and typically-developing children learn words from a directional cue of an arrow versus a potentially intentional referential gaze cue.

The following sections review the literature on word learning with referential gaze and the evidence supporting whether children with ASD treat referential gaze as intentional. First, I will briefly review the research on word learning and referential gaze in infancy. Following this, I will discuss variables that have been related to successful word learning in children with ASD (e.g., language levels, IQ). Next, I will present the evidence pertaining to whether children with ASD understand intention from referential gaze in the word learning context. This evidence comes from study designs that 1) compare children with ASD against children with DD, 2) compare referential gaze against control cues when cues are competing, as well as evidence that comes from the types of measures used including 3) visual attention measures, and 4) in-depth learning measures. Finally, I will conclude with a description of the current study and research questions.

Word Learning and Referential Gaze in Infants

The word learning literature in infancy explores multiple variables that contribute to how infants build their early vocabularies. However, I am interested in the specific variable of referential gaze. When infants learn from referential gaze, why do they do this? This review focuses on infant studies (typical development, under 2 years of age) that explore the specific role of referential gaze[2] in fast-mapping paradigms. I excluded studies that explicitly examined referential gaze and emotional expression (e.g., disappointment, glee; Tomasello, Strosberg, & Akhtar, 1996). Though none of these infant studies have compared learning with referential gaze against a control cue when cues were in isolation, all of the studies below have concluded that learning from referential gaze is because infants treat gaze as an intentional cue.

---

[2] The studies discussed below use another's referential gaze, but because they include live interactions, inevitably gaze itself is not the only cue, for example, head movement, body posture. For consistency, I use the term referential gaze.

The fast-mapping literature is largely defined by the first infant studies that recognized the role of a speaker's nonverbal cues (Baldwin, 1993a; 1993b; Baldwin & Baird, 2001). In an original study, Baldwin (1993b) examined fast-mapping in two different conditions that taught a label to 14- to 19-month-old infants: the follow-in condition and the discrepant condition. In the follow-in condition, a speaker provided infants two novel objects to play with. Once infants focused on their own object, the speaker reached for and held the other object. Then, while infants were looking at their own object, the speaker gazed at the infant's object and uttered a label, "*It's a peri*". In this condition, the speaker followed into infants' focus of attention. In contrast, in the discrepant condition, while infants' focus of attention was on their own object, the speaker held and gazed at the speaker's own object while uttering the label. In this discrepant condition, the speaker did not follow into infants' focus of attention and instead required infants to redirect attention to the speaker's object. Immediately after teaching the label, test trials asked infants to identify the object between the two novel objects presented during teaching, "*There's a peri here. Can you point to the peri? Point to the peri*." The infants' selection of an object at test, or *word recognition*, demonstrated whether they mapped the label to the object during teaching.

Baldwin found that only the oldest group of 18- to 19-month-old infants were able to successfully map in the discrepant condition. She concluded that infants shifted their gaze from what they were looking at and learned in the discrepant condition because infants treated the speaker's gaze as intentional. She also claimed that infants' looking behavior suggested an intentional reading of gaze because infants looked significantly more in the discrepant over the follow-in condition on measures of looking at the speaker, at the speaker's toy, and back and forth between the speaker and the speaker's toy, suggesting that infants were sensitive to where the speaker was looking. However, these older infants could have shifted their gaze to follow the speaker because referential gaze effectively directed their attention, and not neccessarily because it was an intentional cue. Moreover, both younger infants, who were not able to learn in the discrepant condition, and older infants exhibited more looking behavior in the discrepant relative to the follow-in condition on all three looking behavior measures. Thus looking behavior measures do not appear to be strong support for an interpretation of intention understanding if this behavior was also seen in younger children who were unable to demonstrate learning in the discrepant condition.

This rich interpretation of referential gaze is seen in other studies. Hollich et al. (2000) established a different type of discrepant condition by testing an infant's ability to follow a speaker's referential gaze to a less interesting object versus an object that was perceptually salient, or more visually interesting. They found that by 24 months, infants would follow referential gaze for word learning to a less interesting object, instead of mapping the label to their own focus of attention, the more perceptually salient object. These authors also concluded that an intentional reading of gaze helped infants to change their own focus from a more interesting object and instead map the label to the less interesting object. Yet as in Baldwin, the speaker's gaze could have simply been an effective cue at re-directing children's attention, or the speaker's gaze could have increased attention on the speaker's object, making the object more salient. Though no studies to date have compared referential gaze against a control for directional information, some studies have examined referential gaze against a control for attentional salience (Baldwin, 1993a; Baldwin et al., 1996; Moore et al., 1999).

In their experiment 1, Moore et al. (1999) pitted referential gaze against attentional salience. Attentional salience was established by illuminating an object while also making it spin. This design differed slightly from Hollich et al. because Moore et al. activated the salience cue at the same time as presenting referential gaze. In Hollich et al., perceptual salience was used to establish a context of word learning, and not used in directly timed competition as an alternative cue to referential gaze. In Moore et al., the mismatch condition revealed the role of referential gaze: at the same time gaze was directed to a static, unmoving object on one side of the room, an object on the other side of the room was activated to light up and spin. If infants successfully mapped the name to the unmoving object versus the illuminated, moving object, this would demonstrate that infants' learning can be guided by the referential gaze cue and not simply because infants map a name to an attentionally salient object. These authors found that 24-month-olds successfully mapped the name to the static object in the mismatch condition, and mapping was successful even though infants were making more head turns to look at the attentionally salient object versus the static object. These authors concluded that 24-month-olds read referential gaze as intentional, and successful mapping was not because of attentional salience. Yet another explanation for these findings could be that attentional salience is stronger or preferred when it is conveyed by gaze direction rather than illumination/spinning. Because cues were in competition, it is unclear if learning would be similarly successful with attentional salience when it is the only cue available. If similar learning is possible with either attentional salience or referential gaze, then learning from referential gaze, a potentially intentional cue, may not be different from an attentionally salient cue.

In sum, results in infancy have interpreted referential gaze to be an intentional cue during word learning. This rich interpretation is based on evidence such as learning in a discrepant condition (Baldwin, 1993b; Hollich et al., 2000), infant looking behavior such as looking at the speaker, the speaker's object, and coordinating attention between the two (Baldwin, 1993b), and infants learning from referential gaze when in competition with a cue of attentional salience

(Moore et al., 1999). However, this evidence could also be consistent with a lean interpretation, such that infants treat gaze as an effective directional cue, which makes an object more salient. In the current study, I directly test whether children treat referential gaze as a cue that directs their attention, or a cue that conveys a speaker's communicative intent. I use a directional cue of an arrow, which attempted to control for both the line-of-regard and attentional salience, the combination of which was not controlled for in prior infant studies. Rather than competing cues, I examined if learning within individual children differed from an arrow or referential gaze when cues were presented separately in two conditions. By examining learning with each cue, results can demonstrate if children can learn with both cues, cannot learn with either cue, or if children can learn with both cues but learning with one cue is different than another.

Moreover, I expand learning measures to not only include the commonly used measure of word recognition, but also in-depth learning measures (Bani Hani et al., 2012; Norbury et al., 2010). Children can be successful on a measure such as word recognition with only partial knowledge of a word, which is that a label is associated with an object. Yet it is also important to examine if children are encoding what they've learned as a word, which has a meaning (i.e., what an object looks like, how it can be used), and can be generalized to other objects similar in kind.

Before reviewing the evidence supporting whether children with ASD treat referential gaze as intentional in the context of word learning, the next section reviews how success in word learning is in part, related to the heterogeneity of characteristics in children with ASD.

Word Learning from Referential Gaze is Heterogeneous in Children with ASD

Baron-Cohen, Baldwin, and Crowson (1997) were the first to use a fast-mapping paradigm to examine word learning abilities in children with ASD. Using the same live teaching protocol as Baldwin (1993b), Baron-Cohen et al. (1997) included children with ASD, with a mean chronological age of 9 years and a mean language age of 2 years; this large gap indicates that these children were severely delayed. Although 14 of the 17 children could learn in the follow-in condition, the majority of children with ASD could not learn in the discrepant condition, with only 5 of 17 children selecting the target object from a bag of six objects. The authors interpreted that children with ASD were unable to learn the label-object association when the object was outside of their own focus of attention because they did not understand the speaker's intention. However, again, there was no control that could account for the fact that children could have had difficulty with the direction of the speaker's gaze, rather than the speaker's intention. This early study suggests that children with ASD have difficulty with learning in a discrepant condition, though the interpretation for why is still unclear.

However, since this study, multiple factors have been related to children's success on word learning tasks, suggesting that not all children with ASD have difficulty with learning in a discrepant condition. Learning in a discrepant condition is important because in contrast to a follow-in condition where a speaker labels what children are attending to, in the discrepant condition, children must attend to where the speaker is looking in order to learn the word successfully. Most prominently, many word learning studies have demonstrated that children with ASD can learn words in a discrepant condition when they do not have as severe delays as those in the aforementioned study by Baron-Cohen et al. (Akechi, Kikuchi, Tojo, & Osanai, 2013; Akechi et al., 2011; Bani Hani et al., 2012; Gillespie-Lynch, Elias, Escudero, Hutman, & Johnson, 2013; Luyster & Lord, 2009; McDuffie, Yoder, & Stone, 2006; McGregor, Rost, Arenas, Farris-Trimble, & Stiles, 2013b; Parish-Morris, Hennon, Hirsh-Pasek, Golinkoff, &

Tager-Flusberg, 2007; Tenenbaum, Amso, Abar, & Sheinkopf, 2014; Venker, Kover, & Weismer, 2016). Factors related to successful word learning in ASD include children's initial language abilities (Bani Hani et al., 2012; McDuffie et al., 2006; Venker et al., 2016), and nonverbal intelligence (Field, 2016; Gillespie-Lynch et al., 2013; Luyster & Lord, 2009). For example, Bani Hani and colleagues (2012) found that children with ASD with higher receptive and expressive language abilities had better task success in a discrepant word learning condition. In addition to child factors, successful learning in the discrepant condition has also been seen because of environmental factors that have been found to support word learning. For example, children with ASD have shown successful word learning in adapted teaching situations with multiple cues such as pointing, multiple labels, or adding motion to the labeled object (Akechi et al., 2011; 2013; Luyster et al., 2009; Walton & Ingersoll, 2013). Thus child and environmental factors can support discrepant word learning in children with ASD, demonstrating that fast-mapping is not always impaired, as might have been inferred from the conclusions by Baron-Cohen et al. (1997).

Yet despite studies that have shown that many children can learn words, it is still unclear how children with ASD treat referential gaze specifically in contexts of word learning. Though the sample in Baron-Cohen et al. represents a subset of individuals with ASD with severe delays, the overall conclusions of impaired intention understanding in individuals with ASD are consistent with conclusions made another word learning study when children with ASD did not have severe delays (Norbury et al. 2010). However, both studies interpreted a lack of intention understanding in children with ASD when they demonstrated weaker performance relative to children without ASD. Thus these studies make the two assumptions that children without ASD do treat gaze as intentional, and children with ASD would not treat gaze as intentional because of their known social impairments. In contrast, two other studies have suggested that individuals with ASD may treat gaze as an intentional cue (Aldaqre et al., 2015; Bani Hani et al., 2012). Yet these latter two studies still assume that gaze is treated as an intentional cue in typically-developing individuals, when another possibility could be that gaze is more simply read for directional information. In the current study, I address these assumptions by comparing how both children with ASD and typically-developing children treat referential gaze in the context of word learning, and comparing whether the same children treat a directional arrow cue differently from a referential gaze cue. Arrow and referential gaze cues were matched as closely as possible on size, color, and duration of movement, thus I interpret any potential differences between cues as evidence that referential gaze is treated differently because it is read as an intentional cue, and not simply a directional one.

Before describing the current study, the following four sections review the different study designs and measures that have been used as evidence to support whether children with ASD do or do not treat gaze as intentional in a word learning context. Only one of the studies below have investigated alternative interpretations to an intentional reading of referential gaze (Field, 2016). The first two sections reviews two different study designs: 1) comparing children with ASD to children with DD, and 2) competing cues. The last two sections reviews two types of measures where the relative performance of children with ASD versus typically-developing children have been interpreted as indicative of intention understanding: 1) measures of visual attention (speaker, object) and 2) measures of children's in-depth learning.

Comparisons Between ASD and Developmental Delay

To determine if using referential gaze is specifically impaired in those with ASD versus other clinical populations, researchers have compared the performance of children with ASD against children with developmental delay (DD), whose delays do not include a diagnosis of social impairment. If children with ASD are impaired relative to children with DD, then a weaker performance in children with ASD may in fact be because of their social impairments, rather than poor language ability and intellectual delays that characterize both groups (see Nadig & Bang, 2016 for a review on methodological considerations of group comparisons, or matching). It is important to note that individuals with ASD make up a heterogeneous population (Kjelgaard & Tager-Flusberg, 2001; Pickles et al., 2014; Tager-Flusberg & Joseph, 2003), thus individuals who are compared to a DD group, are mainly those with intellectual impairment, which make up roughly one half of the ASD population (Charman et al., 2011). Individuals with ASD with intellectual impairment are difficult to compare with typically-developing children of the same age because of their language or intellectual delays.

One methodological strength of the original Baron-Cohen et al. (1997) study was the inclusion of a DD group. This DD group was matched to the ASD group on chronological age and language age (*M* chronological age = 9 years, *M* language age = 2 years). These authors found that both groups were able to learn in the follow-in condition (15/17). Yet in the discrepant condition, 12 out of 17 individuals with DD were able to use a speaker's referential gaze to switch attention to the speaker's object. This difference was significant in comparison to the 5 out of 17 individuals with ASD. Therefore, the success of individuals with DD (without social impairment) suggests that the difficulty in those with ASD was specific to their social impairment, and not a by-product of a large gap between children's chronological and language age. Other studies including a similar ASD sample with low verbal ability have replicated these learning difficulties in children with ASD (Preissler & Carey, 2005; Walton & Ingersoll, 2013), although some have found successful learning in discrepant conditions in children with ASD with low verbal abilities (Benjamin et al., 2015). Yet only one study since Baron-Cohen et al. has also included children with DD to examine word learning in children with ASD (Field, 2016). In contrast to Baron-Cohen et al. (1997), Field (2016) found learning difficulties in both children with ASD and children with DD, which suggests that children with ASD are not the only group with word learning difficulties, and a more general reason than a lack of intention understanding could explain difficulties for both groups.

Field (2016) used an actor in a video who provided referential gaze and labeled one of two objects. As in Baron-Cohen et al. (1997), Field included children with ASD with low language ages relative to their chronological age (*M* language age around 3.5 years, *M* chronological age around 9 years). However, the author also included children with ASD with higher language ages relative to their chronological age (*M* language age around 6.5 years, *M* chronological age around 10 years). The delineation of low and high language (higher language was relative to their chronological age) was determined with a median split and comparison groups included language-matched children with DD and typically-developing children. Yet in contrast to Baron-Cohen et al. (1997), children with DD did not learn from referential gaze. Instead, for both children with ASD and those with DD who had low language, groups performed at chance. Since children with DD also demonstrated difficulty, these findings challenge whether impaired learning from referential gaze is specific to a social impairment in those with ASD. Moreover, both children with ASD or DD with higher language were able to complete the task, supporting other studies that noted that word learning performance is related to language and not a diagnosis of ASD (Akechi et al., 2011; 2013; Bani Hani et al., 2012; Gillespie-Lynch et al., 2013; Luyster & Lord, 2009; McDuffie et al., 2006; McGregor, Rost, Arenas, Farris-Trimble, & Stiles, 2013b; Parish-Morris et al., 2007; Tenenbaum et al., 2014; Venker et al., 2016). These findings suggest that children with severe language delays have word learning difficulties in these paradigms because of a general difficulty with using cues, and not because social impairments specific to children with ASD affected their ability to read referential gaze as an intentional cue.

To examine this hypothesis, Field investigated whether difficulties in word learning could be seen with cues other than referential gaze, such as an illumination cue or an arrow cue. If children with low language continued to demonstrate difficulties with cues other than referential gaze, then this would support a general impairment with using all cues. In these conditions, an actor was still present and uttered the novel label, but instead of providing referential gaze, the illumination cue or the arrow cue indicated the target object. In one condition, a red arrow appeared next to the target object and pointed at it. In another condition, the illumination cue was a red transparent square that appeared over the target object. Results did not fully support a general impairment in both groups of children, because children with low language in both groups were able to use an arrow for word learning; this occurred in both low and higher language groups. Only the DD group, at both language ages, were able to learn with an illumination cue. These findings, therefore, suggest a specific impairment with referential gaze, and not all other cues, in both children with ASD or DD who have low language abilities.

However, a few key points complicate these findings. For one, the experiment with referential gaze and the experiment with illumination or arrow cues were conducted with two slightly different samples. Therefore, it is unclear whether the same children treated referential gaze differently from an illumination cue or an arrow cue. Additionally, there were multiple differences between the referential gaze cue versus the illumination or arrow cue that make it difficult to compare learning. Both illumination and arrow cues were much larger, were placed horizontally directly next to or on the target objects (referential gaze was a diagonal line-of-regard above the target objects), did not include movement, and were a much more salient color of red. These differences make it difficult to isolate exactly what about an arrow or an illumination cue helped children with ASD learn new words, and potentially many of these differences contributed to the more successful learning seen from arrow and illumination cues relative to a referential gaze cue. Finally, Field did not clarify if teaching occurred in a discrepant condition, and no results were provided that could confirm where children's attention was at the time the target was labeled. This lack of clarification makes it difficult to determine whether learning, at times, was in fact due to the cues or if children simply followed their own focus of attention.

In summary, while some children with ASD have demonstrated difficulty with using referential gaze, specifically those who have a severe gap between their chronological and language age, it is unclear whether this difficulty is specific to children with ASD (Field, 2016). If both children with ASD or DD with severe delays have difficulty with a referential gaze cue in a context of word learning, then these difficulties may be related to factors that are present in both groups of children, and not social impairment specifically. The findings of Field (2016) contradict those of Baron-Cohen et al. (1997), whose children with DD also had severe delays. These contradictions may be due to methodological differences such as the use of a discrepant condition or different DD control groups. In both studies, children with DD were of mixed aetiologies, and in some cases unknown (Baron-Cohen et al., 1997), which does not ensure that DD groups across studies were of similar abilities, or which children were without social impairment. Moreover, while comparisons to a DD group answer questions about specificity, this comparison does not provide sufficient evidence that can answer how children with ASD treat referential gaze. While Field (2016) did test if the difficulty with using referential gaze could be seen with other cues, such as an arrow, methodological concerns limit direct comparison between cues.

I address these concerns in multiple ways. First, I focus on the subgroup of children with ASD who do not have severe delays in their language age relative to their chronological age. By including this subgroup, I am able to focus this dissertation on why children with ASD or typically-developing children learn with referential gaze in cases when they do, rather than on why children with ASD or DD have difficulties with referential gaze. Moreover, I compared if within individual children with ASD, learning differed between a directional arrow cue and a referential gaze cue. Both the arrow and the referential gaze cue were matched on multiple variables such as size, color, and movement. This stringent matching better isolates if in the same children, learning is similar or different with either cue, with differences suggesting that referential gaze contributes to word learning because it is treated as intentional.

Referential Gaze versus Competing Cues in Children and Adults with ASD

As noted in the General Introduction, competing cues can demonstrate a preference for referential gaze over another cue, which has been interpreted as an intentional reading of gaze. However, a preference for gaze could be because of reasons other than intention understanding, such as gaze being a stronger cue than the other competing cue. In ASD research, two studies have tested word learning from referential gaze when in competition with a control cue (Aldaqre et al., 2015; Field, 2016), similar to the mismatch paradigm of Moore et al. (1999), and both have interpreted a preference for gaze as intention understanding.

In her experiment 6, Field (2016) included multiple combinations of competing cues (e.g., arrow cue versus illumination cue, as mentioned above) while an actor uttered a label, but only two conditions included referential gaze: one condition with a referential gaze cue versus an arrow cue, and another condition with a referential gaze cue versus an illumination cue. Field found that when the competing cue was an arrow, children with ASD and higher language more often followed the arrow than the actor's referential gaze, though children did not follow gaze significantly more than chance when the competing cue was illumination. However, as noted above, the arrow cue may have been much more salient than the referential gaze cue (e.g., bigger, brighter, closer to the target), which could explain why it was followed for word learning.

These findings are somewhat contradicted by those of Aldaqre et al. (2015) who, in contrast to all other word learning studies with children, is the only study to include adults with ASD (*M* age 36.9 years, range 19 – 61 years) versus adults with typical development (*M* age 32.5 years, range 20 – 53 years). These authors examined word learning when a referential gaze cue competed against a cue of attentional salience (i.e., jiggling object) and selecting the object indicated by gaze was tested against chance in each group. Chance was set at 25% because during test there were four objects: two which had not been seen before and two which were the target and distractor used during labeling. Chance was also set at 50%, which ignored the two objects that had not been seen before and only considered the target and distractor objects seen during labeling. Aldaqre et al. found that adults with typical development learned the word in comparison to chance responding, at both 25% and 50%, which they interpreted as an intentional reading of gaze. In contrast, while adults with ASD demonstrated learning significantly better than 25% chance responding, learning was not significantly different from 50% chance. Based on these findings, these authors concluded that adults with ASD have some understanding of intention though not to the same degree as those with typical development. This interpretation differs from other studies because these authors conclude that intention understanding does exist, to a certain extent, in individuals with ASD. However, it is still unclear if there are explanations, other than intention understanding, for how a referential gaze cue is used to learn in a word learning context.

As in Moore et al. (1999), these studies provide a picture of whether individuals prefer referential gaze over other competing cues, and these studies suggest that a lack of a preference for referential gaze means that children do not treat gaze as intentional. However, gaze may have been difficult for individuals with ASD in both studies because the competing cue was simply stronger than gaze, and not because individuals were unable to understand intent. Moreover, competing cues do not address why referential gaze itself may be an effective cue. In this study I will compare learning from a directional arrow cue versus learning from a referential gaze cue when cues are not competing. By comparing learning from each cue, when cues are matched on size, motion, and physical features, I can examine if learning itself is different between cues, and not just if there is a preference to follow referential gaze relative to another directional cue.

Thus far, I have discussed study designs that have been used to examine whether referential gaze is understood as an intentional cue. In the following two sections, I review the types of measures that have been used as indicators of referential gaze as an intentional cue in a word learning context: visual attention to the target object and the speaker, and in-depth learning. Performance on these measures have been compared between children with ASD and typically-developing children. These studies have assumed that the performance of typically-developing children serves as a baseline for intention understanding, though it is unclear how children treat a referential gaze cue relative to a control cue.

Visual Attention to the Target Object and the Speaker to Support Intentionality

Children's visual attention to the target object and the speaker provide an important picture of how children use another's referential gaze and potentially, another's intentions. However, all studies that have examined children's visual attention in a context of word learning did not have a control for referential gaze, thus cannot address if visual attention measures reflect a pattern of behavior that could be seen with any other directional cue. Word learning studies with individuals with ASD relative to control groups offer some consistent findings on visual attention to the target object, but mixed findings on visual attention to the speaker. Visual attention is now often measured using eye tracking. Though this method was less common during the time of early infant studies, in ASD research, most of the studies below (except for Preissler & Carey, 2005) have examined visual attention using eye tracking.

Most studies found no differences between children with ASD and typically-developing children in their visual attention to the target object (Akechi et al., 2011; 2013; Norbury et al., 2010; Tenenbaum et al., 2014). This means that children with ASD were able to look at an object of referential gaze to the same extent as typically-developing children. However, a few differences have been noted (Aldaqre et al., 2015; Gliga et al., 2012). For example, though Gliga et al. (2012) did not include children with ASD, they examined children at-risk for ASD (i.e., siblings of children with ASD) versus children with low risk for ASD (i.e., no family history of ASD). The group of children at-risk for ASD were further subdivided into those with poor or typical social skills. These authors found that all groups did not differ on the duration of looking at the target while the actress was providing referential gaze and labeling the target. However, when children were asked to identify the target during test, the subgroup of children at-risk for an ASD with poor social skills looked less at the target and pointed at the correct target less relative to the two groups of children with typical social skills (i.e., at-risk children with typical social skills and children with low risk for an ASD). Based on these findings, these authors concluded that intact gaze following abilities do not guarantee intention understanding of referential gaze, which may be impaired in those with poor social skills.

This interpretation of between-group findings suggests that differences in children's social skills explains differences in the duration of looking at and pointing to the target. However, children with poor social skills in Gliga et al. (2012) also had significantly lower language abilities than both groups of children with typical social skills. These lower language abilities suggest that another interpretation of these findings could be that language abilities, social skills, or some interaction of both could be a reason for less looking at and less pointing to the target during test. This example illustrates that the evidence to conclude a lack of intention understanding in children with ASD (or those at-risk for ASD) is limited when based solely on between-groups differences, which relies on assuming that groups are similar on at least, all known factors, except for differences in social skills.

In addition to visual attention to a target object, attention to a speaker's face/eyes has been suggested as an indicator of intention understanding, though it is unclear whether children would show this same attention to a cue other than referential gaze. In the original Baldwin (1993b) study, part of the evidence used to support an interpretation of intention understanding was that infants looked more at the speaker, the speaker's toy and back and forth between the speaker and the speaker's toy in the discrepant relative to the follow-in condition. Similar measures have been used in children with ASD relative to typically-developing children, and studies have interpreted that when children with ASD look less to the speaker's face that this is evidence to support a lack of intention understanding (Norbury et al., 2010; Preissler & Carey, 2005). However, first it is unclear if children with typical development would attend similarly or differently to a control cue that was not considered intentional versus a potentially intentional cue of a speaker's gaze. Second, it is unclear if children with ASD would always attend less relative to typically-developing children whether the cue was gaze or another cue.

Others have not found differences between children with ASD and typically-developing children in their attention to the speaker (Akechi et al., 2011; 2013; Tenenbaum et al., 2014), though the role of intention understanding cannot be determined without a control for referential gaze. One reason for these mixed findings could be that attention to a speaker may be related to variation in children's language abilities (Falck-Ytter, Fernell, Hedvall, Hofsten, & Gillberg, 2012). Studies that have found differences included children with low language ages (Preissler & Carey, 2005), while studies that have found similarities between groups of children included those with age appropriate language abilities (Akechi et al., 2011; 2013). These findings suggest that less attention to a speaker in children with ASD relative to control groups may, in part, be explained by factors other than intention understanding of referential gaze.

Additionally, while the aforementioned studies have examined attention to the speaker's face or head turn, only one study has examined attention to the specific area of the speaker's eyes (Tenenbaum et al., 2014). The question of the specific role of a speaker's referential gaze has not been carefully examined in the context of word learning. Though word learning studies have discussed the role of referential gaze, these studies often include multiple cues, among which, referential gaze is noted as a particularly important cue. Therefore, it is important to clarify whether it is referential gaze itself that may contribute to intention understanding in this context. Only Tenenbaum et al. (2014) measured visual attention specifically around the area of the speaker's eyes, and found that children with ASD looked at the area of the speaker's eyes for similar durations as control groups (e.g., children with language delay and typically-developing children) in a word learning context. Moreover, during test, these authors did not find differences between groups in how quickly they located the target (i.e., latency). In exploratory analyses, these authors found that after accounting for age and IQ scores, only in children with ASD, longer looking to the speaker's eyes during labeling was significantly associated with faster latencies at test. These findings demonstrate that in children with ASD, the variation in their attention to a speaker's eyes was related to how quickly they located the target object, which suggests that they were able to use this cue to aid their word learning. However, there was no control for the speaker's eyes, which makes it difficult to determine if attention to the speaker's eyes is indicative of intention understanding and whether it was this intention understanding that helped children to more quickly locate the target during test.

Thus, while some results of visual attention to the target object and the speaker have exhibited less attention in children with ASD relative to control groups, others have found no differences between groups. Some studies have interpreted less visual attention in children with ASD relative to control groups as evidence of a lack of intention understanding. However, for all groups of children it is critical to include a control cue for referential gaze. By comparing visual attention between referential gaze relative to a cue that controls for the direction of attention, the findings can demonstrate whether, in the context of word learning, visual attention patterns to the target and the speaker is unique to referential gaze or can be seen with another directional cue. Moreover, the connection between following referential gaze and intention understanding is also tenuous when many studies have only focused on attention to the speaker's face, and few have examined children's attention to the specific area around the speaker's eyes (Tenenbaum et al., 2014).

To better understand if attention to the target object and attention to the speaker can indicate intention understanding, I compared both between- and within-children with ASD and typically-developing children on measures of visual attention. These measures include: attention to the target object and attention to the area of the speaker's gaze, as well as how often children look back and forth from the speaker's gaze to the target object. Within groups comparisons examined how children in both groups attended to a simple directional cue of an arrow in contrast to a referential gaze cue, where any within group differences could be attributed to an intentional reading of referential gaze. I use an area of interest around the speaker's eyes relative to an area of interest of the same size around an arrow cue, to allow for a fair comparison when cues are matched on the size of the area.

In-depth Learning to Support Intentionality

Finally, to examine how referential gaze contributes to multiple aspects of learning, I extended measures to include those that can demonstrate in-depth learning, beyond the association between a label and single object. A word is a symbol that people use to communicate about an object or a concept. To demonstrate a more complete understanding of the meaning of that object or concept, measures need to go more in-depth. In this dissertation in-depth knowledge was defined as examining learning beyond word recognition, and includes measures that assess children's semantic knowledge (e.g., what the object looks like, how the object can be used), and how children generalize the word. Two word learning studies with children with ASD have used in-depth learning measures that could speak to the lack of (Norbury et al., 2010), or presence of (Bani Hani et al., 2012) attributing intent to referential gaze. However, because neither study included a control cue for referential gaze, both studies have assumed that gaze contributes to learning because it is an intentional cue. It is not known if similar in-depth learning would be seen with a control cue as well as gaze.

In a word learning paradigm that differs from the prior fast-mapping paradigms, Norbury et al. (2010) compared children with ASD relative to typically-developing children matched on chronological age, nonverbal IQ, and receptive vocabulary, though children significantly differed on a standardized measure of expressive description abilities. The same four novel labels were presented in each of the two conditions, three times in each condition. In the neutral condition, an actress in a video provided direct gaze to a participant in the presence of three objects, and in the biased condition, an actress provided referential gaze to one of three objects. In both conditions, after a short baseline period, children were asked to, "*Show me the [novel label]*", and children used a computer mouse to select one of the three objects. Attention to the speaker and the target was measured with visual attention measures and learning was measured with word recognition, word description, and word production immediately after teaching and four weeks later. This study is novel in using multiple measures of word learning and assessing word learning over time in children with ASD.

Relative to typically-developing children, children with ASD provided less visual attention to the actress (see section on visual attention above), which the authors interpreted as a less developed understanding of gaze as an intentional cue. On the measure of word production, children with ASD were significantly better than typically-developing children at producing a greater proportion of phonemes of the target label when asked to name pictures of the objects, although typically-developing children caught up to children with ASD four weeks later. Yet the most striking finding from this study was that during the description task, children with ASD recalled a smaller proportion of semantic features (e.g., colors and shape that describe the target). This difference was more pronounced four weeks later, although the interaction between group and visit did not reach significance ($p = .08$). However, children with ASD were impaired relative to typically-developing children on a standardized measure of expressive description abilities, thus it could be that these children with ASD had difficulties with describing objects in general.

In contrast to Norbury et al., other studies have not found differences between children with ASD and control groups (i.e., typically-developing children and/or children with specific language impairment) in the semantic information recalled with familiar words (McGregor et al., 2012) or when teaching novel words (Gladfelter, 2014; Kreger, 2016). Thus, it may be that children with ASD do not have impaired descriptions relative to control groups. However, these latter studies did not include referential gaze when teaching novel words. The findings of Norbury et al. are interesting because they suggest that there may be a link between when children use an intentional reading of gaze to learn new words and their recollection of semantic features. One way to test this relationship is to compare children's descriptions when they are taught novel words with a non-intentional control cue or a potentially intentional cue of referential gaze. It could be possible that both children with ASD and typically-developing children describe more with a referential gaze cue versus a control cue, but perhaps children with ASD still describe less than typically-developing children. This latter scenario would suggest that both groups of children may read gaze as an intentional cue, but that children with ASD do have difficulty with recalling semantic features relative to typically-developing children.

One study challenges the view that children with ASD do not understand intent in a word learning context. Bani Hani, Gonzalez-Barrero and Nadig (2012) assessed word learning using measures of word recognition and word generalization in children with ASD or typical development. The generalization measure assessed children's abilities to generalize the learned label to other objects that are similar in kind, or other exemplars of the object (i.e., generalizing from a 3D object to black and white images). The authors found no differences between groups (see Benjamin et al., 2015 for similar findings with children with ASD), and proposed that successful generalization may be because children with ASD understood the ostensive cues used in the interaction (Csibra & Gergely, 2009), although it is unclear from this study if there are other reasons for generalization. Bani Hani et al. based their proposal on the work of Csibra and Gergely (2009), in which the authors state that human ostensive cues, such as direct gaze (often used before a speaker's referential gaze), have been evolutionarily selected from human communication to be perceived as intentional by infants. Csiba and Gergely suggest that these ostensive cues signal to infants that they must pay attention and that infants will then expect to learn something that they encode as potentially generalizable. Yet in Bani Hani et al. (2012), if both groups could have also generalized using a cue that is not considered as intentional, then learning could be due to other factors and not an intentional reading of referential gaze.

Therefore, though there are studies that suggest that children with ASD can or cannot use referential gaze as an intentional cue for in-depth word learning, there is no evidence to support that intention understanding is the only reason for in-depth word learning. Between-group comparisons can demonstrate if word learning is as robust in each group, yet this method is limited at addressing whether children with ASD and typically-developing children treat referential gaze as an intentional cue. To test if in-depth learning occurs from intention understanding, I compared in-depth learning within individual children when learning from a directional arrow cue versus a referential gaze cue. With the stringent matching between cues (on size, motion, physical features), in both groups of children, better in-depth learning with a referential gaze cue relative to an arrow cue would indicate that intention understanding may benefit individual children's word learning.

I included in-depth learning measures that assess children's understanding of the meaning of newly learned words. These included prior measures such as *word description* and *word generalization*. Moreover, I included a previously unused measure of semantic knowledge in fast-mapping literature, *word associations*. In the word association task children were asked to provide the first word they can think of when they hear a target word. Word association tasks are often used to assess an individual's semantic network, or how words and word meanings are interconnected (e.g., *dog* primes *cat* through a semantic association; Collins & Loftus, 1975; Quillian, 1969; Sheng & McGregor, 2010). If children are encoding word meaning, then I would expect to see children provide responses that are semantically related to the word (e.g., what an object looks like and how it can be used). Therefore, I included a word association task to examine if it could be a different measure to examine children's semantic knowledge of new words. Finally, thus far, only one study to date has tested word learning over time (Norbury et al., 2010), which demonstrates children's retention and consolidation of newly learned words. I added to this literature by examining children's learning both immediately after teaching and one week later.

The Current Study

I investigated whether 6- to 11-year-old children with ASD or typical development use referential gaze, and potentially, another's intentions in the context of word learning. Learning was compared between an arrow cue, a control for the direction of attention, and a referential gaze cue. Any potential differences between cues would be interpreted as indicating that children may be treating referential gaze as an intentional cue. This age range was selected to be comparable to the ages of children in prior studies of word learning when word descriptions were required (McGregor et al., 2012; Norbury et al., 2010), and children in studies of action prediction (Vivanti et al., 2011). Children with ASD and typically-developing children were matched on nonverbal IQ and chronological age, with nonverbal IQ within a normal range. Therefore, it should be noted that this study represents only a subgroup of the full heterogeneity of the autism spectrum. After matching, children did not differ on other demographic variables such as the ratio of girls to boys and parent level of education.

Children watched a video where they were taught a label-object association by either an arrow cue or a referential gaze cue using a fast-mapping paradigm. In contrast to prior studies, all other nonverbal cues were avoided (e.g., head movement, body movement, voice direction), thus allowing for the interpretation to be specific to referential gaze itself for word learning. Importantly, I tried to establish a discrepant learning condition similar to that of prior studies in word learning (Baldwin, 1993b; Baron-Cohen et al., 1997). A discrepant situation is important to ensure that children's learning is a result of following a cue, and not because children were already interested in an object. I adapted a paradigm similar to Hollich et al. (2000), where the conflict was established using perceptual salience. Thus, at the outset of the video, the distractor object (i.e., the non-target object) was made to be more visually interesting to look at and either a referential gaze or an arrow cue was directed to the less interesting target object. Though perceptual salience does not guarantee a discrepant learning situation for all children, all target and distractor objects were paired based on results of pilot testing to better ensure that children would be drawn to distractors over targets (see Appendix B for more details on pilot testing).

Thus this discrepant situation demonstrates a stronger test of children following a cue, because children would need to shift their own focus of attention from a perceptually salient object, the distractor, to the less salient target object.

In addition to the teaching and test phases seen in prior studies, I included a baseline phase during the videos. Baseline phases were included to examine children's initial interests in the target and distractor objects before teaching and test phases. Therefore, I was able to examine if children followed cues to look at the target more during teaching or test phases relative to their initial looking during the baseline phase. Using an eye tracker, I examined attention to the area of the target object relative to the distractor object, and attention to the area of the cue (i.e., an arrow or referential gaze) across all phases.

Word learning was assessed immediately after teaching and one week later using measures that have been suggested to be indicative of intention understanding in prior word learning studies. These measures included: 1) *word description* (children described the object), (2) *word generalization* (children decided if images of the objects that varied from the original, such as by color and shape, were exemplars of the learned object), and (3) *word production* (children provided the name of the object when shown a picture of the object). I also included a previously unused measure in fast-mapping literature, *word association* (children provided the first word they could think of when they heard the target word).

This study advances our fundamental understanding of whether children with ASD potentially treat another's gaze as an intentional cue, and links this fundamental understanding to real-world implications of children's in-depth learning of new words. This understanding is important because although studies have proposed that intention understanding is impaired or preserved in individuals with ASD, no studies to date have explored alternative explanations for how children treat referential gaze in a context of word learning. I examine one potential explanation, which is whether referential gaze is treated as a directional cue, rather than an intentional one. Moreover, I extend this study to address real-world implications of learning by examining measures of in-depth learning and the retention of learning over time (i.e., one week later). One potential clinical application of this study involves learning from in-depth word knowledge measures. For example, if children with ASD do have difficulty with intention understanding and this results in weaker word descriptions, then we can more precisely address what additional information needs to be stressed to better support word meaning in children with ASD.

Research Questions and Predictions

Because children in both groups were well-matched, I predicted few differences between children with ASD and typically developing children (e.g., Akechi et al., 2011; 2013; Bani Hani et al., 2012; Luyster & Lord, 2009). Instead, I predicted within-group differences in responses to cue conditions in both groups of children. I argue that differences between an arrow cue and a referential gaze cue are evidence of children treating referential gaze as an intentional cue, and not simply a cue that directs attention. I predicted differences between arrow and gaze conditions on all measures, particularly those that have been suggested to be indicative of another's intent: children's attention to gaze (Baldwin, 1993b; Norbury et al., 2010; Preissler & Carey, 2005), word recognition (Baron-Cohen et al., 1997), word description (Norbury et al., 2010), and word generalization (Bani Hani et al., 2012). For simplicity, referential gaze will hereby be referred to as gaze. Below is a list of the research questions for the current study on the measures of visual attention and learning.

1. **Visual attention to the target object: Do children with ASD and typically-developing children follow cues and locate the target during test? Do children attend to the target object differently between cue conditions?** I measured looking time to the target object across baseline, teaching, and test phases.
2. **Visual attention to the cue: Do children with ASD and typically-developing children attend differently to a gaze versus an arrow cue?** I measured looking time to the cue area in baseline and teaching phases and the number of looks between the cue area and the target object (i.e., contingent looks) in the teaching phase.
3. **Visual attention to the target object during test: Does latency to the target differ between gaze and arrow conditions in children with ASD and typically-developing children?** I measured the latency to first look at the target object during test (Tenenbaum et al., 2014).
4. **Learning: Do children with ASD and typically-developing children demonstrate better word recognition with gaze versus an arrow cue?** I measured children's correct pointing to the target object during test.
5. **In-depth learning: Do children with ASD and typically-developing children demonstrate better in-depth learning immediately after the video and one week later with a gaze versus an arrow cue?** I measured word association, word description, word generalization, and word production both immediately after the video and one week later. I also looked descriptively at the consistency of in-depth learning across measures at both visits.

METHODS

Participants

This study was approved by the McGill University Institutional Review Board. Participants included 6- to 11-year-old children with ASD and typically-developing (TD) children recruited in Montreal, Canada. Participants were English-speaking or French-speaking, defined by parent identification of their child's dominant language. Parents were asked to report the approximate percent per week of children's language exposure in the home from birth to the child's current grade. An average of these percentages, reflecting lifetime exposure, verified children were exposed to their dominant language the majority of the time at home for English-speaking ASD (M = 93.59%, SD = 10.39%), English-speaking TD (M = 78.75%, SD = 19.93%), French-speaking ASD (M = 70.81%, SD = 32.48%), and French-speaking TD children (M = 80.54%, SD = 21.53%). Parents provided informed consent and children provided informed assent prior to study participation. Children received a small prize for their participant at each visit.

Twenty-seven children with ASD were recruited from multiple sources. These sources included: a list of previous participants in our lab who agreed to be contacted for future studies, a Quebec government social service organization serving children with disabilities, special schools for children with disabilities in the Montreal area, community events for families of children with autism, a university research database, and flyers distributed to organizations that work with children with disabilities. Two children were dropped from the study because 1 child was not able to complete the study protocol, and 1 child did not meet criteria for ASD on the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2002) or the Social

Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003). This resulted in 25 children with ASD in our final sample.

Clinical diagnosis of ASD was established through a multidisciplinary assessment. All children had a documented diagnosis of ASD under the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria, reflecting subtypes in the DSM-IV criteria (American Psychiatric Association, 2013). For 20 of the 25 children, this multidisciplinary assessment included the ADOS. For the other 5 children, the multidisciplinary assessment did not mention use of the ADOS per se, but other social interaction activities similar to those seen in the ADOS. When the ADOS was administered, this was done by a clinical psychologist or psychiatrist in a multidisciplinary team. Autism diagnoses were confirmed upon study entry with the SCQ, a parent report screener for ASD. Although the standard cut-off score of 15 or above is used to identify individuals with ASD, we adopted a cut-off score of 12 or above on the SCQ which has been shown to be more sensitive in identifying younger children with ASD (Allen, Silove, Williams, & Hutchins, 2007; Corsello et al., 2007; Wiggins, Bakeman, Adamson, & Robins, 2007). Children with ASD did not have any other medical conditions associated with ASD (e.g., fragile X syndrome) and no physical vision or hearing limitations that would interfere with study procedures (e.g., color blindness). Except for 1 child with strabismus who was not included in eye tracking analyses, all other children had normal or corrected vision by parent report, with 2 children who wore glasses. Eight children with ASD were diagnosed with comorbidities, which included attention-deficit/hyperactivity disorder (ADHD), speech dyspraxia or language impairment. Four children with ASD were on medication for attention at the time of testing.

Forty-seven TD children were recruited from the same university research database and flyers distributed in the Montreal area. Four children were not included in the analyses because they participated in pilot versions of the experiment. Our final sample included 43 TD children. TD children did not have any developmental learning or behavior disorder by parent report and did not meet criteria for ASD on the SCQ. Additionally, children did not have 1st or 2nd degree relatives with ASD (accounting for potential hereditary characteristics), or any physical, vison or hearing limitations that would interfere with study procedures (e.g., color blindness). All but 1 child had normal or corrected vision by parent report, with 6 children who wore glasses. This 1 child was reported to have corrected vision, but had a history of strabismus, thus was not included in eye tracking analyses.

Matched groups. The current study is a quasi-experimental design, meaning that participants in each group of children are not assigned randomly, but are determined because of a pre-existing characteristic that determined group assignment (e.g., presence of ASD). Because of this pre-existing characteristic, studies must match groups of children on all other known characteristics (e.g., IQ, age) to best interpret any between-group differences as being attributable to the pre-existing characteristic of interest (e.g., the presence of ASD). There are many different methods to match groups in quasi-experimental studies (Kover & Atwood, 2013). In the current sample, children with ASD and TD children were matched using a statistical method known as propensity scores. Propensity scores are calculated for each child and represent an individual's probability of being in a group given the specified covariates (Rosenbaum & Rubin, 1983). The specified covariates are determined from prior research theory (Kover & Atwood, 2013; Stuart, 2010), and propensity scores can be calculated from logistic regression.

Though there are multiple benefits to using propensity scores, I used this method for two main reasons. First, propensity scores avoid the subjective bias of hand selecting matches when the person matching children may know about children's behavior and study performance. Second, propensity scores can summarize multiple covariates into a single scalar score, which avoids the difficulty of matching closely on every covariate (Stuart & Rubin, 2008). More information of benefits can be seen in Appendix A. Notwithstanding these benefits, some limitations of using propensity scores include their limited use in research in neurodevelopmental disabilities (Blackford, 2007; see Blackford, 2009 for use in children with Down Syndrome), and sample size requirements. Although some authors suggest large sample sizes are required (Kover & Atwood, 2013), Blackford (2007) suggest 5 – 10 participants per covariate in the main group of interest (e.g., children with ASD) to be sufficient.

For this sample, I specified nonverbal IQ and age as covariates in the propensity scores. Following Blackford's guidelines (2007), I included only two covariates based on my main group of interest, my sample of 25 ASD participants (approximating 10 participants for each covariate). Nonverbal IQ and age were selected because of their known relationships with referential gaze following and with language outcomes in children with ASD (Falck-Ytter et al., 2012; Leekam, Hunnisett, & Moore, 1998; Thurm, Lord, Lee, & Newschaffer, 2007). Though parental education and gender are also known for their relationships with language (Fenson et al., 2007; Hoff, 2006), IQ and age were prioritized because of their known relationships in children with ASD. Language ability was not included as a covariate because distributions of the primary language ability measure (i.e., scaled scores CELF-4 Word Classes) revealed inherent differences between groups that would result in excluding too many children with ASD when the goal was to retain as many children with ASD as possible; children with ASD ranged from 2 standard deviations below to 2 standard deviations above the normal range on scaled scores of the CELF-4 Word Classes.

Propensity scores and the final matched sample was determined using MatchIt, a program written for R that calculates propensity scores and provides a list of children in the matched group (Ho, Imai, King, & Stuart, 2011). Full detail of the matching procedure can be seen in Appendix A. The final matched group included 24 children with ASD and 24 TD children, which removed 1 child with ASD, who I determined to be an outlier based on visual inspection of the propensity score distribution. A 1 to 1 matching was used with the *optimal* method in MatchIt. A 1 to 1 matching includes the same number of children with ASD and TD children. The optimal method reduces the difference in propensity scores between matched pairs while taking into account the overall set of matches (Stuart, 2010). Although propensity scores are a tool to aid matching, the final sample needs to be verified regarding whether groups are well matched. Whether the groups were well matched was verified using the standardized mean difference (using the standard deviation in the treatment group as the denominator) and variance ratio of propensity scores between groups (Stuart, 2010). Recommended guidelines for well-matched propensity scores include a standardized mean difference less than 0.25 and variance ratios close to 1 (i.e., within 0.5 to 2; Rubin, 2001). The matched sample in this study had a standardized mean difference of .15 and variance ratio of 1.04 for propensity scores, indicating that propensity scores including IQ and age were balanced between groups.

Table 1 exhibits characteristics on the matched sample. Because propensity scores are relatively new to ASD research, individual characteristics were also evaluated with more familiar measures. Continuous variables were evaluated with visual inspection of boxplots, Cohen's *d*, variance ratios, and results of paired sample *t*-tests. Categorical variables were evaluated with

Fisher's exact test. Guidelines to evaluate well matched groups include Cohen's *d* close to 0, variance ratios close to 1, and *p* values > .5 (Cohen, 1988; Kover & Atwood, 2013; Mervis & Robinson, 1999; Mervis & Klein-Tasman, 2004). Cohen's *d* was calculated using the compute.es package (Del Re, 2013) in R version 3.3.0 (R Core Team, 2016) with formulas in line with Kover and Atwood (2013).

In summary, groups were balanced on their propensity scores, as well as all individual covariates collected, except for language abilities. The individual covariates of IQ (d = -.09, variance ratio = .91, p = .75) and age (d = .11, variance ratio = 1.27, p = .71) also satisfied criteria for balanced groups. Groups did not significantly differ on characteristics of parental education and the ratio of boys to girls, although neither meet the strict matching guidelines of p > .5 (Mervis & Robinson, 1999; Mervis & Klein-Tasman, 2004). Groups significantly differed on their distribution of language abilities as measured by the CELF-4, with Cohen's *d* effect sizes greater than or equal to .25 and variance ratios greater than or equal to 1.49. As expected and consistent with their diagnosis, all children with ASD demonstrated significantly higher scores on the SCQ and significantly lower scores on the socialization subscale of the Vineland Adaptive Behavior Scales, Second Edition (VABS-II; Sparrow, Cicchetti, & Balla, 2005). There were also similar proportions of English-speaking and French-speaking children in both groups of children. Finally, the number of children who had different orders of the cue condition and the context presentation were balanced in each group (King & Nielsen, 2016).

Table 1. Matched Sample Characteristics: 24 Children with ASD and 24 TD Children

	ASD	TD	*p*	*d*	vr
Age[a]	8.83 (1.26) 6.67 – 11.33	8.70 (1.12) 6.50 – 10.50	.72	.11	1.27
Nonverbal IQ[a]	108.29 (12.65) 87 – 131	109.50 (13.24) 83 – 131	.76	-.09	.91
CELF-4 Word Classes Total[a^]	9.74 (3.74) 2 – 16	12.08 (3.06) 6 – 19	.02*	-.69	1.49
CELF-4 Recalling Sentences[a]	8.08 (4.16) 1 – 14	11.17 (2.18) 7 – 16	00**	-.93	3.64
CELF-4 Word Associations[a]	29.96 (15.02) 5 – 65	33.29 (11.17) 17 – 53	.38	-.25	1.81
Socialization subscale - Vineland[a]	76.76 (11.67) 61 – 110	110.00 (11.88) 80 – 129	.00***	-2.82	0.96
Social Communication Questionnaire[a]	20.88 (5.82) 12 – 32	4.42 (2.62) 0 – 9	.00***	3.64	4.95
Gender (M : F)	21 : 3	18 : 6	.46		
Maternal education (below : above university)[#]	12 : 12	6 : 18	.14		
Number of English and French children (En : Fr)	11 : 13	10 : 14	1		

Continuous and categorical variables were analyzed using paired sample *t*-tests and Fisher's exact tests, respectively. *p* = *p* value from significance test, *d* = Cohen's *d*, vr = variance ratio. Negative values for Cohen's *d* indicate higher values in the TD group.
[a]The values shown are the mean (*SD*) and range. * $p < .05$, two tailed. ** $p < .01$, two tailed, ***$p < .001$, two tailed.
[^]One child with ASD did not complete this measure thus the *p* value was calculated from an independent samples *t*-test rather than a paired *t*-test.
[#]For one TD child the father's education was used instead of the mother's education because the mother's education was not provided.

Background Measures

Language abilities were measured using subtests of the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003) or the validated version for Canadian French speakers, the Évaluation clinique des notions langagières fondamentales – version pour francophones du Canada (Secord, Wiig, Boulianne, Semel, & Labelle, 2009). The CELF-4 is a standardized measure of receptive and expressive language abilities comprised of multiple subtests to evaluate different aspects of children's language. Subtests can be combined to represent different domains of language such as Core Language or areas of Receptive and Expressive language. However, due to time constraints, I only included subtests which evaluated aspects of children's language most closely related to the experimental measures. These included subtests that reflected children's semantic language rather than other subtests that measured children's ability to follow directions or morphosyntax. The semantic language measures used included the *Word Classes* and *Word Associations* subtests. I also included the *Recalling Sentences* subtest to evaluate children's verbal working memory and distinguish children with or without language impairment (Conti-Ramsden, Botting, & Faragher, 2001; Norbury et al., 2009). Normative scores are available for each subtest, with a mean scaled score of 10 and a standard deviation of 3.

In *Word Classes,* children were asked to verbally identify and describe relationships between words that share a semantic relationship. For example, a child is verbally presented with four words: fish, milk, fin, and spider, and then asked to provide the two words that are related (i.e., *fish and fin*) as a measure of receptive language/comprehension. After children identify the word pair, they are asked to describe how the two words go together as a measure of expressive language (i.e., *fish have fins, or fish use their fins to swim*). Norms are available separately for receptive and expressive language, as well as a combined score, the Word Classes – Total score. The Total score was used in this study because outcomes involved both children's receptive and expressive language abilities. Furthermore, receptive and expressive language scores on this measure were highly correlated within each group (ASD $r = .83$, TD $r = .81$, ps $< .001$), suggesting that a combined score would be an adequate representation of both receptive and expressive language abilities.

In *Word Associations*, children were asked to list as many category members as they can in 1 minute for the categories of animals, food, and jobs/occupations, which assess children's semantic category knowledge. In *Recalling Sentences*, children were asked to repeat sentences of increasing length and syntactic complexity. Sentence repetition tasks have been found to have high sensitivity and specificity to detect language impairment in children, as well as assess verbal working memory abilities (Archibald & Joanisse, 2009; Conti-Ramsden et al., 2001). In the current sample of children, 6 children with ASD and 1 TD child met standard research criterion of language impairment based on 1 standard deviation below the mean scaled score on the Recalling Sentences subtest (Archibald & Joanisse, 2009; Conti-Ramsden et al., 2001). Of these children, all 5 children with ASD and no TD children were also below 2 standard deviations. For the typically-developing child who was below 1 standard deviation, the parent did not report any developmental concerns, thus the child was retained in this study.

Nonverbal IQ was assessed with the Leiter International Performance Scale, Third Edition (Leiter-3; Roid & Miller, 2013). The Leiter-3 is a standardized measure of children's visualization and reasoning skills across four subtests: Figure Ground, Form Completion, Classification and Analogies, and Sequential Order. All four subtests are incorporated into a standard score with a mean of 100 and standard deviation of 15. Administration of the Leiter-3 and children's responses are conducted without language. Instead, experimenters use simple gestures (e.g., pointing) to demonstrate how to complete the tasks and children respond by pointing or manipulating cards or blocks.

Children's social skills were assessed with the Vineland Adaptive Behavior Scales, Second Edition – Socialization domain on the Parent/Caregiver Rating Form (VABS-II). The VABS-II is a parent report of multiple domains of children's adaptive behavior including Communication, Daily Living Skills, Socialization, Motor Skills, and Maladaptive Behavior. The domain of interest for the present study was the Socialization domain, which provides a measure of children's everyday social behavior. The Socialization domain is comprised of three subsections of Interpersonal Relationships, Play and Leisure Time, and Coping Skills. All three subsections are incorporated into a standard score with a mean of 100 and standard deviation of 15.

Parents filled out questionnaires regarding their level of education (as a proxy for socioeconomic status). As discussed above, parents also filled out a questionnaire of children's history of language exposure across their lifetime in different settings of home, school, and other environments, though only the home setting was used as noted above. The home setting was used over school and other environments because the lifetime exposure in the home setting more closely matched children's dominant language provided by parent report.

ASD confirmation/exclusion. The SCQ is a 40-item parent questionnaire that screens for symptoms of ASD. Conventionally a score of 15 or above meets criteria for individuals with ASD. However, we adopted a cut-off score of 12 or above on the SCQ which has been shown to be more sensitive in identifying younger children with ASD (Allen et al., 2007; Corsello et al., 2007; Wiggins et al., 2007). Though specificity scores have been reported to be lower with this cut-off (Corsello et al., 2007), meaning that some TD children may be falsely classified with ASD, no TD children in our sample were falsely classified. In the ASD sample, 21 children had scores of 15 or higher, 2 children had a score of 14 and 1 child had a score of 12 on the SCQ. All TD children had scores of 11 or below.

Testing Protocol

Testing was conducted over the course of two visits, approximately one week apart. At visit 1, parents filled out the VABS-II, SCQ, and demographic information while the child watched videos and completed other standardized assessments in a separate room. During the videos, the experimenter stood behind an opaque curtain where they could not see the video to remain blind to the cue condition. Cue conditions were identical in their auditory stimuli, thus the experimenter could not have known a cue condition from listening to the videos. I was in a separate room from the videos where I controlled the video and could view the child's face on a separate monitor. I took notes on whether the child was attending to the video, in case there were any concerns regarding the eye tracking data that needed to be reviewed. For one child, I was the experimenter while another experimenter controlled the video. For this child, I was blind to the cue condition.

Visit 1 lasted approximately 1.5 – 2 hours and the progression of activities often occurred as follows (please see the Stimuli section below for more detail regarding the trials in each block): (1) CELF-4 Word Associations, (2) block 1 trials, (3) CELF-4 Recalling Sentences, (4) 5- to 10-minute break, (6) block 2 trials, (7) CELF-4 Word Classes. Visit 2 lasted approximately 1 – 1.5 hours and the progression of activities often occurred as follows: (1) block 1 trials (no videos, only in-depth learning measures) (2) Leiter-3, and (3) block 2 trials (no videos, only in-depth learning measures). At times, the order of standardized assessments was adapted to children's interests, though block 1 and block 2 trials were always separated by a break. If children participated in a prior study in our lab within the past year, standardized assessments (e.g., CELF measures, Leiter) were not re-administered to avoid potential test-retest effects and fatigue, as suggested by the authors of the assessments (Roid & Miller, 2013; Semel et al., 2003). In the matched sample, 15 children with ASD and 5 TD children had previously participated in prior studies in our lab. Though visit 1 was always conducted in our lab testing room, visit 2 was either at the lab or at the families' homes depending on the families' preferences.

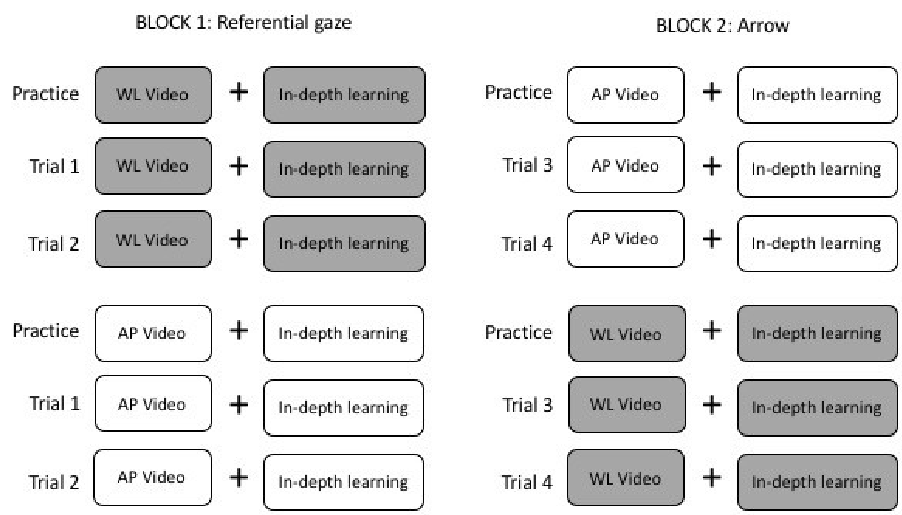
Apparatus

A remote faceLAB eye tracker (version 4.5.1) recorded children's eye movements at a rate of 60Hz. The eye tracker uses an infrared light to measure the corneal reflection from participants' eyes to the screen, and calculates the location and duration of children's eye movements. Video stimuli were presented using GazeTracker presentation software (version 8.0.3156.1000 FULL for faceLAB) on a 43-inch TV monitor (diagonal distance) approximately 93 cm from the child. The infrared light was approximately 62 cm from the child. Videos were shown at a resolution of 1280 x 720 units. Fixations were defined using a 40 pixel spatial and 100 ms temporal parameter (Gliga et al., 2012, Vivanti et al., 2011).

Stimuli

Block design. As seen in Figure 1, children were presented with word learning videos in two different conditions, where the target was indicated with either a gaze or an arrow cue, presented in separate blocks; block order was counterbalanced across children. There was one practice trial and two trials per cue condition. In addition to word learning videos, children viewed action prediction videos (discussed in Chapter 2). For each child, context order presentation was reversed such that the context order in block 2 was reversed from the order in block 1. Cue and context order presentation was counterbalanced across children. The final distribution of the cue and context order can be seen in Table 2.

Figure 1. Example of Study Protocol Presented to One Participant



WL = word learning, AP = action prediction. In-depth word learning measures were presented in the following order: 1) word association, 2) word description, and 3) word generalization. In-depth learning measures of action prediction were presented in the following order: 1) original set and 2) different color set.

Table 2. Matched Sample: Number of Children in Different Cue and Context Order Presentations

	ASD			TD		
	Gaze first	Arrow first	Total	Gaze first	Arrow first	Total
Word learning first	8	5	13	6	5	11
Action prediction first	4	7	11	7	6	13
Total	12	12	24	13	11	24

Cue comparison. Because the goal of this study was to evaluate the specific role of gaze, videos did not include head or body motion. Instead, the actor simply provided direct gaze, with gaze facing the participant, then shifted her eyes to an object that was on either side of her head. Although restricting head and body motion resulted in less naturalistic teaching, this allowed for an evaluation of the effects of referential gaze alone.

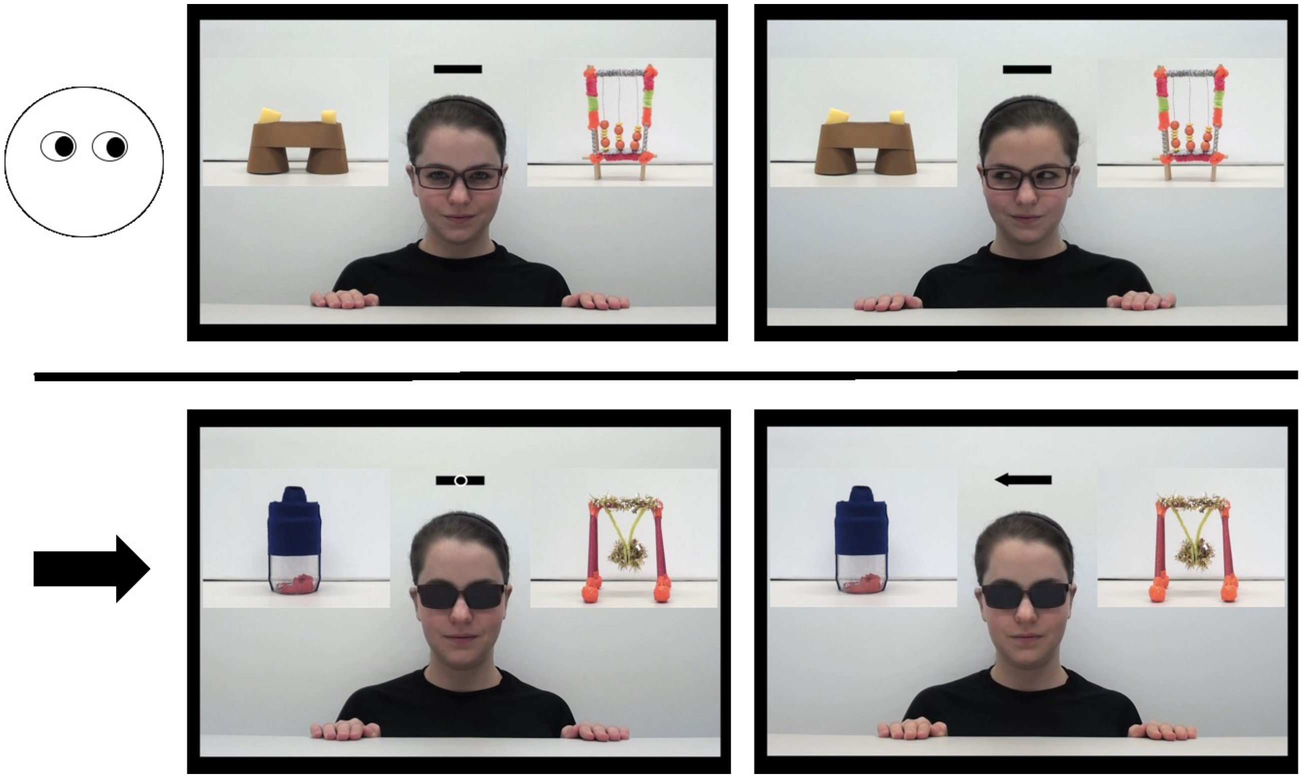
An arrow cue is a commonly employed control for referential gaze in the literature (e.g., Birmingham et al., 2012; Senju et al., 2004) and matches as a control for gaze for multiple reasons. First, arrows are likely familiar cues to children 6- to 11-years old, and can be seen in everyday life from street signs to school activities. Although familiarity between gaze and an arrow cue is not equivalent since children are exposed to gaze from birth, an arrow is likely to be a familiar directional cues to children at this age. Second, both can serve as a sign to convey direction, in addition to increasing salience of the referent. Third, an arrow is not only familiar and a sign of direction, but it is a conventional cue, in which we have learned that this cue directs attention like gaze (Hommel, Pratt, Colzato, & Godijn, 2001). In addition, using an arrow cue extends the large body of spatial cueing literature where an arrow cue is used.

Using an arrow also allowed for critical features to be matched between the two cue conditions including the presence of the actor, and matching both cues on their size, approximate shape, and duration of motion. As seen in Figure 2, videos for both cues included the same actor in the center wearing glasses. In the gaze condition, the actor wore clear glasses and a black bar was present above her head. In the arrow condition, the actor wore darkened glasses and a black bar was present above her head that included a white outline of a circle in the center of the bar. During teaching, the actor in the gaze condition shifted her eyes horizontally to the target. In the arrow condition, the circle moved horizontally and turned into an arrow to point to the target.

Therefore, the angle of the direction was horizontal in both, and the movement duration of the circle was matched to the gaze shift duration (10 frames). The size of the arrow was the same height (bottom of eyes to top of eyelid) and width of the eyes (the left corner of the left eye to the right corner of the right eye, from the participant's point of view). The diameter of the circle was similar to that of the opened eyes. As seen in Figure 2, the white outline of the circle was designed to mimic the salience of gaze, suggested to be due to the contrast of the white sclera and the iris (Emery, 2000; Kobayashi & Kohshima, 1997).

Although the same actor was used in the gaze and arrow conditions, the two conditions needed to be filmed separately, because it was too difficult to digitally overlay the darkened glasses to the same video as in the referential gaze condition. The final video files chosen in each condition were matched as closely as possible to have a similar hairline and neutral mouth expression. The actor's mouth was intended to be slightly positive, because during stimuli creation a more neutral expression was found to be too off-putting.

Figure 2. Close up of Gaze and Arrow Cues



The top row above the horizontal line depicts the gaze condition and the bottom row depicts the arrow condition. The left side depicts direct gaze (top) and the white outline of the circle designed to mimic the contrast of the white sclera and the iris (bottom). The right side depicts the gaze (top) and arrow condition (bottom) after the cues have moved, which results in referential gaze to one side and an arrow pointing to one side.

Novel labels. Novel labels were created in a prior norming study in our lab (Howarth, 2010). A subset of four novel labels were selected for this study and further details regarding novel label selection and the audio stimuli can be seen in Appendix B. Novel words and phrases (e.g., *Where is the pagoune*, *Now point to the pagoune*) were recorded in English and French by a bilingual English and Quebec-French speaker in a sound proof booth using a Marantz PMD660 recorder.

Novel objects. Four target objects and four distractor objects were designed for this study. Target objects were designed to be visually less interesting than distractor objects by having limited features and decoration, such as only two colors per target, and simple shapes and materials (e.g., triangle, paper). I used limited features that would likely be simple to describe for children this age. Distractor objects were designed to be visually more interesting, and perceptually salient objects by being shiny, multi-colored, and more detailed than target objects (e.g., multiple little beads, gemstones), thus creating a discrepant condition similar to prior word learning studies (Hollich et al., 2000; Parish-Morris et al., 2007). Therefore, correctly mapping the novel word to the target object would demonstrate that children could change their focus of attention from the distractor object to the focus of attention provide by the cue, the target object, better ensuring that children noticed the cue.

Each target and distractor object had a unique cause and effect function and all objects were roughly similar in size. Careful consideration was taken to create target objects that could be uniquely distinguished from other target objects and distractor objects such that children could demonstrate learning of specific target objects on in-depth learning measures. Based on adult pilot testing, every target object was given a novel label and a fixed distractor object. Details about object creation and adult pilot testing can be seen in Appendix B. There were two trials per cue condition, thus all children were taught four target-distractor pairings, two pairings per cue condition. Pairings were not fixed to cue conditions, and were counterbalanced such that all pairings were shown with either the gaze or the arrow cue across children.

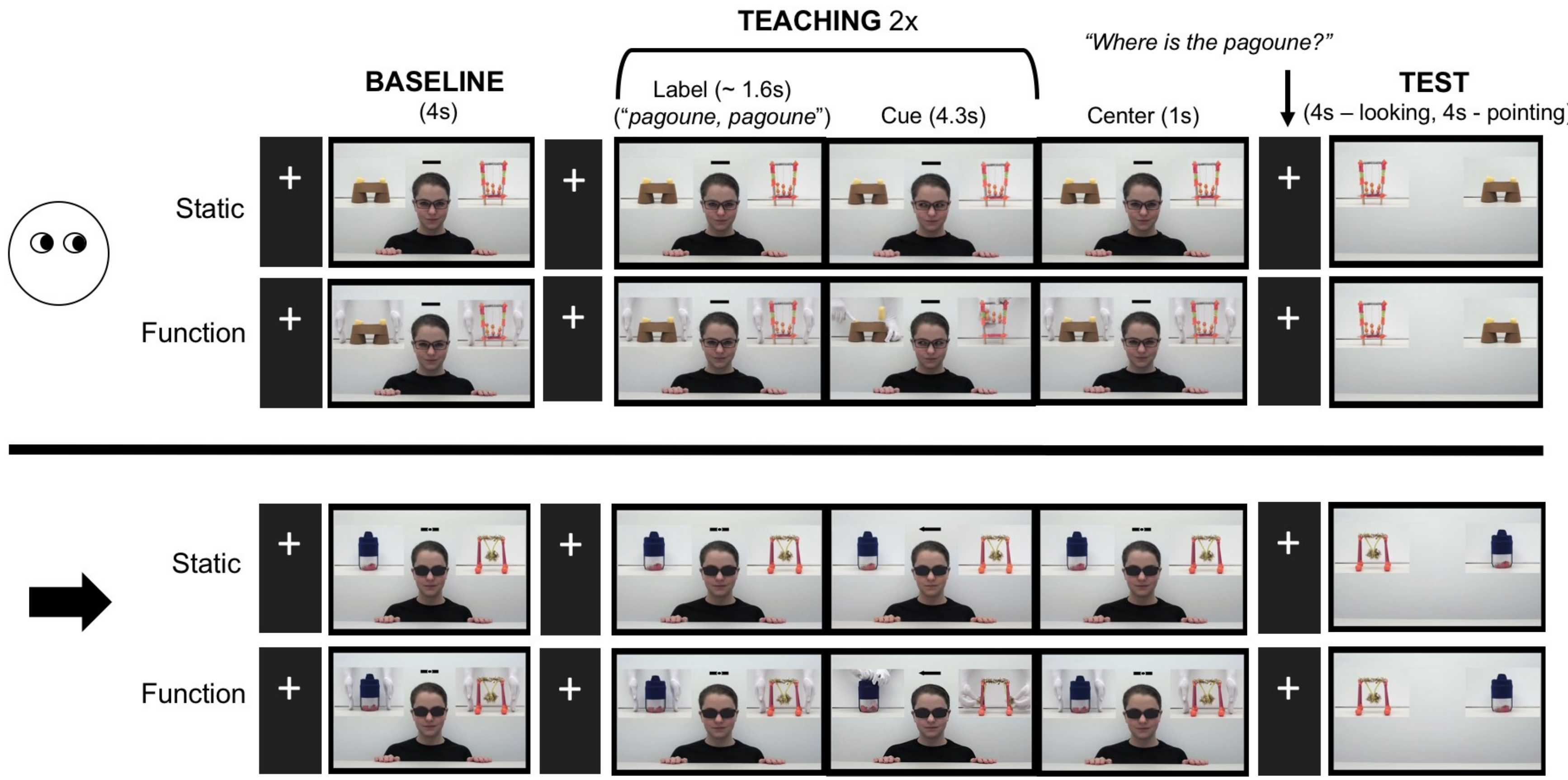
Videos. Videos were created using Final Cut Pro software (version 6.0.6). For each cue condition, children received one practice trial and two experimental trials, for a total of two practice trials and four experimental trials. The number of experimental trials is similar to that of other studies in word learning (e.g., Baron-Cohen et al. 1997; Parish-Morris et al., 2007). A trial refers to both the video and in-depth learning measures presented after the video.

Practice trials consisted of objects familiar to children in this age range: hammer, cup, scissors and glasses (hammer-glasses and scissors-cup were fixed target-distractor object pairs in block 1 and block 2, respectively, though the experimental object order did change). Familiar objects were selected from a list of early acquired household names on the MacArthur-Bates Communicative Development Inventories (Fenson et al., 2007). The purpose of the practice trial was to confirm that children understood the procedure of watching videos, pointing to the correct object, and answering questions about what they saw. No explicit feedback of whether the child pointed correctly was provided for any practice or experiment videos, although experimenters did provide motivational phrases such as "*Great!*" For each cue condition, practice trials were administered before experiment videos and children were told, "*You're going to watch a video and learn a word. You might already know this word. I want you to pay close attention because I'm going to ask you questions about what you learn after the video.*" All children demonstrated that they could pass the practice trials before watching experiment videos. A pass in the practice video was defined by correctly pointing to the familiar object.

Before watching experiment videos, children were told, "*Now you'll learn some new words for things that you haven't seen before. Make sure you pay close attention because I'm going to ask you questions about what you learn after the video like we did before [with the practice videos].*" Children watched the videos shown in Figure 3, first with static objects (top row) and then immediately after with hands that demonstrated object functions (bottom row). Only visual attention to the static videos, and not the function videos, will be analyzed for this study, because static videos are more consistent with prior studies. Function videos included an additional naming episode and were intended to provide additional information about the objects beyond their physical appearance (i.e., what they could do). Interstimulus intervals (ISI) were intervals that separated different parts of the videos and lasted 1.5 s. **1) Baseline:** the child could view the scene for 4s. **2) Teaching:** the novel label was spoken twice, "*pagoune, pagoune*" (~ 1.6 s) while the actor provided direct gaze (gaze condition) or the actor was wearing glasses and there was a static circle above the actor's head (arrow condition); the portion of time when children heard the label is the *label portion* of the teaching phase. All verbal cues were recorded by the same female actor in the videos, though the actor's mouth did not move in the videos. Next, the cue moved to indicate the target object for 3.6 s, then returned to center; from the moment the cue began to move to the moment the cue returned to center is considered the *cue portion* of the teaching phase. The duration of the cue portion totaled to 4.3 s, because it included the cue shift to the target, the cue indicating the target, then the cue shifting back to center. The same teaching phase was presented twice. **3) Test:** prior to seeing the objects during test, children heard the prompt, *"Where is the pagoune?"* during the ISI, then the objects were presented for 4 s. The end of this phrase was set to the last frame of the ISI, meaning children saw the test images immediately after the prompt. Target object side was switched at test to better dissociate children having learned the label was associated with the target object and not the side the cue indicated. After 4 s of viewing the objects, children had another 4 s to point to the target, "*Now point to the pagoune*" for a measure of their explicit word recognition, rather than looking time. Durations of each phase were based on examining what felt natural when making the videos, and similar durations used in prior studies (Gliga et al., 2012).

Whereas in many prior studies a speaker uttered the label while simultaneously providing the referential gaze cue, the label was presented differently in this study. In this study, the speaker never uttered the label herself though the child heard the label, and then the referential gaze cue or the arrow cue indicated the target. I dissociated the label from the cue because children may have thought the speaker was aware of the cue if they heard the speaker label the object at the same time the cue indicated the target. I did not want children to potentially think that the speaker was aware of the arrow cue, in case this might make the arrow seem related to the speaker and potentially intentional.

Figure 3. Word Learning: Frames from the Video that Depict an Example of a Word Learning Video Sequence



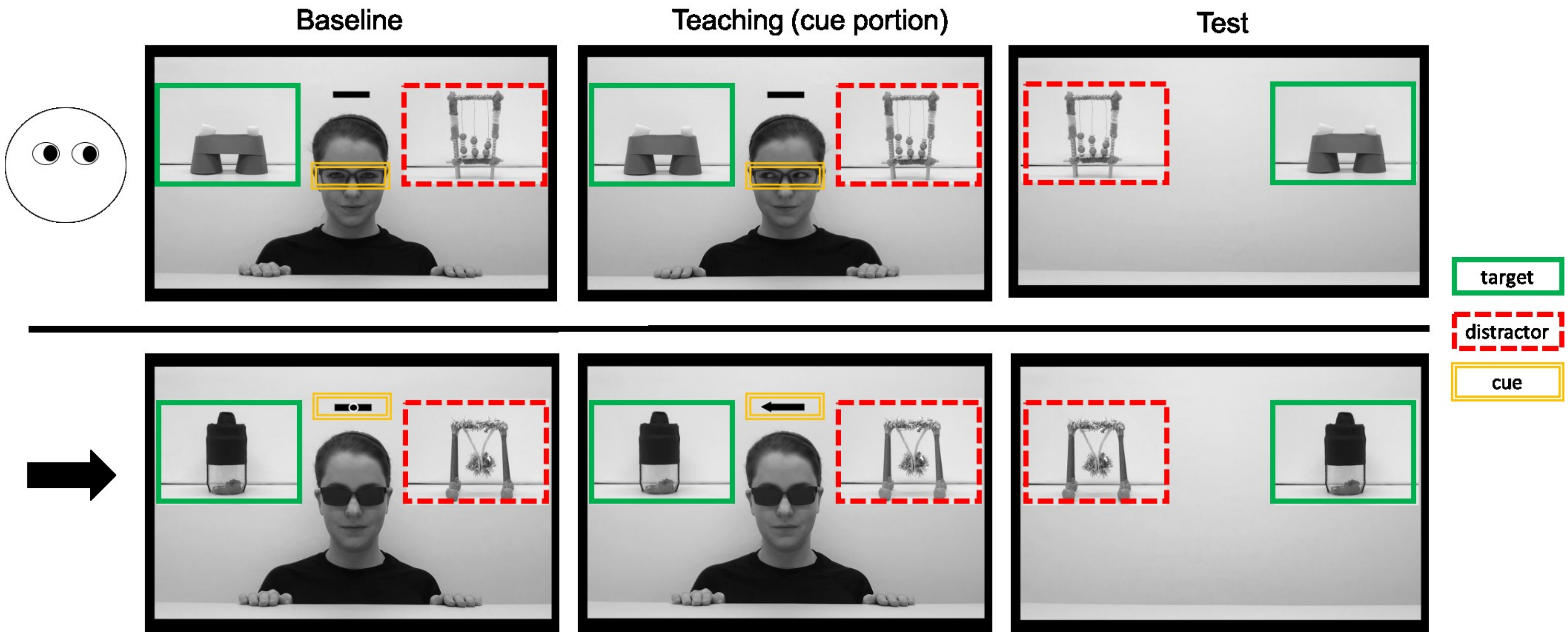
Frames from the referential gaze condition (above the solid line) and the arrow cue condition (below the solid line). Children viewed function videos immediately following static videos in both cue conditions. The duration of the video for the combined static and function video sequence was approximately 1 minute per target object. Children had two experimental trials per cue condition (with different target and distractor objects).

Eye Tracking Video Stimuli

The three video phases used in this study included: 1) baseline, 2) teaching, and 3) test. For analysis, data from the teaching phase included only the cue portion (see the Video section above for details), which was presented twice for each target object. Looking time when the child heard the label (label portion prior to the cue portion) was not included in the current study.

As seen in Figure 4, there were three areas were of interest in each of the video phases: 1) target object 2) distractor object 3) cue area (not included during the test phase). The areas around the target and distractor objects were the same size (width 440 pixels and height 306 pixels), and the areas around the gaze and the arrow were the same size (width 184 pixels and height 54 pixels). The area around the gaze was based on Xu and Tanaka (2013), which included the eyebrows and some space below the eyes. The area of the arrow was matched to the area of the gaze, and the area of the arrow was placed directly above the area of gaze so that cues were in the same central location though the area of the arrow was higher than the area of gaze.

Figure 4. Word Learning: Areas of Interest During the Baseline, Teaching, and Test Phase



To more clearly highlight the areas of interest, the video frames are shown in grayscale, though the videos were shown in color as can be seen in Figure 1. The target object is highlighted in the green solid line, the distractor object in the red dotted line, and the cue in a light orange with double lines.

In-depth Learning Measures

Word association. The word association task measured the strength of the lexical-semantic organization of the target and familiar words (Sheng & McGregor, 2010), and was the first task administered after the children saw the video. We asked children, "*Now back to our word game. I want you to tell me the first word you think of when I name something. Ready?*" For each target word, two familiar words were also presented as part of the task to keep children motivated in case the target words were too difficult. Children provided one response for each target word and each familiar word.

Children's responses were coded into one of six groups: semantic, clang, distractor, any experimental object, error, and other. Semantic associations were defined as responses that indicated children understood the meaning of the object, which in the case of this study, referred to what the object looked like and how it was used. Thus semantic associations referred to intended functions (*e.g., pagoune – pop*), physical attributes (*e.g., pagoune – yellow*), what the object resembles (*e.g., pagoune – binoculars*), real world knowledge for familiar objects only (*e.g., hammer – builder*) and category associations for familiar objects only (*e.g., hammer – tool*). In contrast to associations with familiar words that can be more easily be judged by a coder, semantic associations of target words were more difficult to identify as specific to a particular target word. Thus, for responses to be coded as semantic, they needed to be a response that was considered as reasonably specific to the target object viewed in that trial versus a response that could relate to any object (e.g., *paper*). The coder followed the decision tree seen in Appendix D to find the best coding group.

Clang associations were words that alliterated (*e.g., pagoune – penguin*) or shared similar sounds with the target words (*e.g., pagoune – lagoon, pagoune - p*), and could not be considered as a semantic relationship as defined above. *Distractor* referred to responses specific to the distractor presented with that target object (*e.g., pagoune – beads,* because the distractor object for the pagoune had beads). The category of *any experimental object* included responses deemed too general to be coded as semantic because the response could describe another target or distractor object (*e.g, pagoune – paper*). *Errors* included "I don't know", inflections (*e.g., pagoune – pagounes*), and repetitions (*e.g., pagoune – pagoune*). The category of *other* included responses that bore no clear relationship to the target word (*e.g., pagoune – music*).

Word description. After the word association task, the word description task measured the semantic features children recalled from the word learning episode (Norbury et al., 2010). We asked children, "*Now I want you to describe a pagoune for my friend. Remember you can tell me about the size, color, shape, what you can do with it, and what kind of object it is. Can you tell me three things about a pagoune?*" Children were prompted to provide three more things and the experimenter stopped when the child said that they could not think of anymore. Asking for "*three more things"* was meant to provide a tangible number to children, in the event that saying *"anything"* would be abstract and potentially difficult, thus children did not have to provide three things with this prompt. In some cases, it was clear that children were not offering new information but did not want to say that they did not know anymore, thus ending the task was left to the experimenter's discretion. Experimenters transcribed descriptions (using words, not phonemic or phonetic transcription) from a video recording and coding was completed in two passes.

In the first pass, the coder identified each object based on the description. Because children were not always describing the target object, descriptions needed to be identified for those that described the target object versus those that did not. These identifications better ensured that when semantic features of the objects were counted, that these features actually reflected target objects that were taught. When the coder could identify the target from children's descriptions, these were considered valid descriptions. All other descriptions were considered invalid descriptions of the target, though the coder still attempted to identify what object the child was describing. A description was considered valid when the child provided a minimum of either a reference to the specific shape of the target (e.g., *looks like binoculars* for a pagoune), one color of the target that could not be confused with the distractor (e.g., *yellow* for a pagoune because there was no yellow on the distractor), or both one reference to the cause or effect of the object's function and one other reference to the object's shape or physical attributes (e.g., *pop* (effect of function), *brown* (physical attribute) for a pagoune). A decision tree can be seen in Appendix E and was used by the coder to facilitate object identifications. Descriptions were considered invalid when instead of the target, they described the distractor, a different target, a real object that was not the target (e.g., describing an igloo instead of a pagoune), or the description was too general to determine what the child described. Table 3 includes examples of different children's descriptions when asked to describe the *pagoune*, and how the coder identified each description.

After the first pass of object identification, in the second pass, the coder counted the number of features, or details about the object, in 11 possible coding groups: 1) intended function, 2) physical attributes, 3) non-specific features, 4) imagined function, 5) object category, 6) resemblance, 7) real world facts, 8) invented 9) name comment 10) other 11) repetition. As seen above in word associations, the meaning of an object in this study was defined by what the object looked like and how it was used, thus a variable of *semantic features* was operationally defined by only two categories: *intended function* and *physical attributes*[3]. Though the category of *resemblance* could be specific to the target object (e.g., *pagoune looks like binoculars*), this coding group was excluded for two main reasons. First, not all of the target objects resembled familiar objects to the same degree based on children's responses, meaning that some target objects might have increased semantic features more than others if this coding group was included. Additionally, this coding group was ambiguous regarding how to count the features. For example, if children said a *pagoune looks like binoculars* (counted as one feature for *looks like binoculars*), but another child described the pagoune as being *two cylinders and a rectangle* (counted as 3 features; two, cylinder, rectangle), the latter description would be considered as providing more information though the first child's description of the pagoune looking like binoculars provides similar information, since binoculars have two lenses and are connected in the middle by something that is shaped like a rectangle. Explanations for why additional categories were excluded can be seen below.

The way the 11 coding groups were defined was adapted from prior studies (Nadig, Vivanti, & Ozonoff, 2009; Norbury et al., 2010). *Intended functions* included features that described the specific use or purpose of the object (e.g., *press a pagoune*). *Physical attributes* included features regarding the color, shape, number, relative length or weight of object parts (e.g., simply saying the object is large/heavy was not deemed specific enough to be coded here, the child needed to provide more information about the relative length, weight, or size, such as, *the buttons are small* versus saying *it is small* would be coded in non-specific features), and

---

[3] Regardless of whether the description was identified as a valid description of the target or not, all descriptions were coded for semantic features, though semantic features of invalid descriptions were not analyzed in the current study.

spatial location of physical properties (e.g., *top, bottom, center, attached*). Physical attributes of familiar and target objects were coded slightly differently because of children's known experiences with the objects (e.g., saying that a hammer was heavy was counted as a physical attribute, whereas weight was not considered for target objects since children did not have any physical experience with the objects). In the few instances where physical attributes were referenced with idiosyncratic terminology (e.g., referring to the yellow buttons of the pagoune as pineapples), these were counted towards the category of physical attributes since children did not know what the objects were made of.

Table 3. Examples of Different Object Identifications and Descriptions of the Pagoune

Object Identification by Coder	Excerpts of Descriptions (from different children)
Correct Target (Valid)	*"It's brown, it's small, and you can push down one side then the other side comes up…there's two buttons that are yellow. It looks like it's made out of paper and that's it."*
Distractor (Invalid)	*"on the edge, on the side there are feathers…it's made out of wood, there's like strings and there's like balls on the strings that go up and down…"*
General (Invalid)	*"It's for making sound effects, it's for people to play with and that means it's a toy."*
Other – Igloo (Invalid)	*"you run in it, it's a nice house…it's made by blocks of ice, chunks of really hard ice…"*

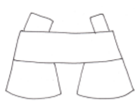
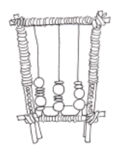
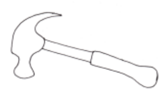
This table provides examples of different children's descriptions when asked to describe the pagoune. A picture of the pagoune can be seen in Figure 5. The column on the left includes how the coder identified the objects in each description. The valid description is the first description, which correctly describes the pagoune. The invalid descriptions were provided in response to asking children to describe a pagoune.

Additional coding groups, along with the group of resemblance, were excluded from the dependent variable of semantic features, because features in the excluded coding groups could not be linked to a specific target object, were made up and not consistent with what was seen in the video, or were repetitions within the same description (e.g., children said *a pagoune is brown* twice). *Real word facts* (e.g., *a pagoune is hand-made*) were excluded because though in some cases children were correct, this was not information they were taught thus could not be known to every child. *Object category* (e.g., *a pagoune is a toy*) was excluded because it was unclear if all children were considering target objects to be in the same object category (e.g., one child mentioned that the objects were for an office while another child mentioned it was a toy). *Imagined function* consisted of made up features about how a child thought the object functioned, but could not be verified from the video (e.g., a child refers to a lever on the side of the pagoune, even though from the video the child never sees a lever). *Non-specific features* were excluded because features in this coding group were too general (e.g., *it is small, it can be any color*). Finally, the coding groups of *name comment* (e.g., *pagoune like lagoon*), *other* (e.g., the pagoune was on the left side of the screen), and *invented* features (e.g., the object is made in a steel factory) were excluded because they were not considered as providing any conceptual information regarding the meaning of the word (e.g., what it looks like, how it can be used).

Word generalization. After the word description task, the word generalization task measured the extension of the target label to other exemplars of the object (Bani Hani et al., 2012; Benjamin et al., 2015). Images of the target, distractor, and familiar objects were shown using E-prime (version 2.0.10.353) on a Dell laptop computer and children were asked to push a red or a green button, "*Now we're going to play a game where you're going to see some objects. Some of them might be pagounes. If you think it's a pagoune, press the green button. If you don't think it's a pagoune press the red button. Are you ready? Let's play!*" Experimenters asked children, *"Is this is a pagoune?/What about this one?"* throughout the 11 images. The accuracy of extensions and response times were recorded.

As seen in Figure 5, for each target object, children were shown six images that consisted of the target object and its exemplars. These exemplars included: (1) the same image seen in the video, (2) a black and white image, (3) an image of the target object that was made using different colors, (4) an image of the target object that was made using different shapes and colors, but was still constructed to perform a similar function (though the function was never shown to the child), (5) a line drawing of the whole object, and (6) a line drawing of the object that deconstructed the different parts of the object. In addition to the target and its exemplars, five non-target objects were presented which included the distractor and familiar objects. The images of the distractor object were shown to see if the child mis-mapped the novel label to the distractor during teaching. Familiar objects (e.g., hammer, bottle) were shown to ensure children were paying attention. In each target object trial, children viewed 11 images in total that were presented in a randomized order.

Figure 5. Images for Word Generalization of Pagoune

	Target (pagoune)	Distractor	Familiar
Original Color			
Black and White			-
Color		-	-
Shape and Color		-	-
Line Drawing of Whole Object			
Line Drawing in Parts		-	-

The figure above depicts exemplars of the target, distractor, and familiar objects. The original colors of the pagoune are brown for the base and rectangle across the top, and yellow for the pieces sticking out on each side. For the pagoune, the different color exemplar is pink for the base and rectangle across the top, with blue pieces sticking out on top, and the different shape and color exemplar has orange pieces for the base and rectangle across the top, with dark green pieces sticking out on top.

Word production. The word production task measured children's ability to verbally recall the target label (Norbury et al., 2010). Because production often comes after comprehension in word learning (Fenson et al., 2007), the word production task was considered as another in-depth learning measure that goes beyond asking children to select an object. Children were shown one image at a time in randomized order, for a total of eight images in their original color: the four target and four distractor objects. Image presentation was randomized at each visit. Children were told, "*You are going to see some pictures of objects and some of them have names, and some of them do not. If it has a name, I want you to tell me the name when I show you the object.*" For each picture children were asked if the object has a name/if they could tell the experimenter what the object was called. Responses to the targets were coded into coding groups of correct/partial productions or incorrect productions. Correct productions included fully correct labels (e.g., *pagoune*) or partially correct because the response included similar phonemes (e.g., *paloule* or *spagoune* for pagoune). Incorrect productions included: 1) child says that he/she doesn't know, 2) the child says the object does not have a name when it was a target object, 3) real names that try to describe the target object (e.g., *balance beam* for pagoune) , 4) nonsense names (e.g., *tochi* for pagoune), 5) the wrong name for the target (e.g., the child says *mimole* when the pagoune was shown), 6) unclear because the phonemes combined more than one label (e.g., *mopen* could be a combination of fopam and mimole), and 7) repetition, because children repeated the same name for two or more objects. Additionally, responses for distractor images were also coded for whether the child mis-mapped the full target label or partial target label to the distractor.

An overview of all measures analyzed in this task, as well as the phases they were collected in, is presented in Figure 6. In-depth learning measures were presented in the following order for both visits: 1) word association, 2) word description, 3) word generalization, and 4) word production. The order of the measures was set to minimize the influence of one measure to the next. For example, because word generalization included pictures of all the objects, this measure could not be before word association and word description. Additionally, I was concerned that word description may prime effects on the word association measure, thus word description was administered after word association to limit priming effects. Although measures (1) – (3) were all presented in succession after the video, word production (4) was not presented with the other measures because children may simply provide the label they were hearing multiple times in the video and during in-depth learning measures. Therefore, word production for all words was assessed as the last measure in both visits.

Coding training and reliability. All measures except for word generalization were open-ended and responses were evaluated through coding systems. One coder was assigned to word association and word description, and a second was assigned to word production. With each coder on their respective measures, I created a coding protocol based on prior research (Nadig et al., 2009; Sheng & McGregor, 2010). Creating the coding protocol was a dynamic process that included defining coding groups based on prior research, applying these coding groups to a subset of existing data, and changing the definition of these coding groups until both coders were in agreement that the coding groups were clearly defined for subsequent use. For each measure, after a coding protocol was created, reliability was achieved between myself and each coder. The coder and I each independently coded data for 13 participants and if the reliability statistics, either kappa or intraclass correlations (ICC), were above .80 for coding groups that were central to research questions, then all data for remaining participants were coded by the coder. The value of .80 was based on recommended cut-offs used for kappa and ICCs in the literature, where .80 and above is considered a strong level of agreement (Hallgren, 2012). If reliability statistics were below .80 for coding groups central to research questions, then the coder and I met to revise the coding protocol based on where we had disagreements, then the coder and I each independently coded another set of 13 participants.

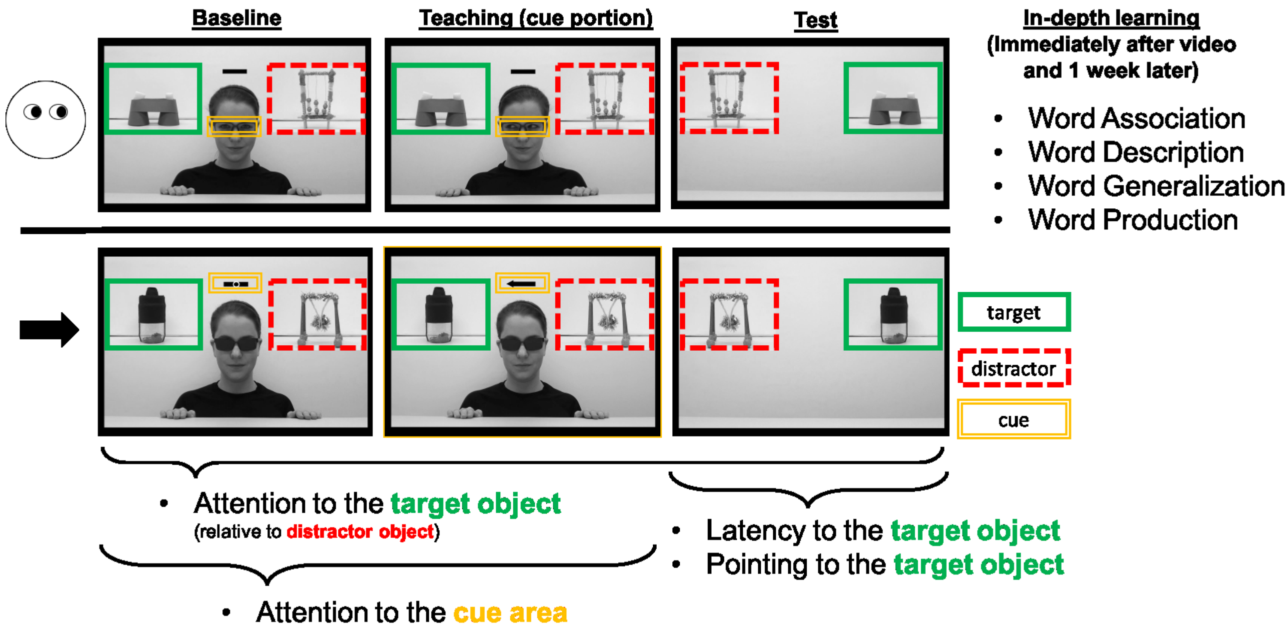
This procedure of independently coding 13 participants, checking reliability, and revising the coding protocol continued until reliability statistics were above .80 for coding groups central to research questions. For word description, this procedure was followed twice for object identification, and only once for the number of object features. For word association, this procedure was followed three times, and for word production this procedure was followed only once. Coders were also experimenters who tested children, but were blind to study hypotheses and cue conditions while testing and coding. I was blind to cue condition during coding. Kappas were used for categorical variables of word association, object identification for word description, and word production. ICCs were used for continuous variables of the number of object features for word descriptions. Kappas and ICCs were calculated in R with the psych (Revelle, 2016) and irr packages (Gamer, Lemon, Fellows, & Puspendra, 2012) and can be seen in Table 4. Kappas and ICCs were .84 and above for the coding groups analyzed in this study, demonstrating that coders were able to achieve a strong level of agreement.

Table 4. Kappa and Intraclass Coefficient Statistics and Confidence Intervals for Reliability

Measure	Variable	Kappa or ICC	CI
Word association	Overall word associations	.91	[.87, .95]
	Familiar word associations	.95	[.90, 1]
	Target word associations	**.84**	**[.76, .92]**
Word description	**Object Identifications**	**.92**	**[.86, .98]**
	Intended function	**.93**	**[.91, .95]**
	Physical attributes	**.95**	**[.94, .96]**
	Non-specific features	.81	[.74, .86]
	Imagined function	.55	[.44, .65]
	Category	.91	[.87, .93]
	Resemblance	.83	[.77, .87]
	Real world	.89	[.86, .92]
	Invented	.47	[.34, .59]
	Name comment	1	[1, 1]
	Other	.55	[.43, .65]
	Repetition	.81	[.75, .86]
Word production	**Word production**	**.92**	**[.88, .96]**

Coding groups in bold and underlined are those that were analyzed in this study. All other coding groups are included here to demonstrate reliability on all groups, but were not analyzed further in this study because they were not central to the research questions. For word description, the coding groups of imagined function, invented, and other were lower in reliability, but this may have been due to the low frequency of descriptions in these groups.

Figure 6. Word Learning: Overview of Measures Analyzed



To more clearly highlight the areas of interest, the video frames are shown in grayscale, though the videos were shown in color as can be seen in Figure 1. The target object is highlighted in the green solid line, the distractor object in the red dotted line, and the cue in a light orange with double lines.

Measures of *attention to the target object* were collected during baseline, teaching, and test phases, *attention to the cue area* was collected during baseline and/or teaching phases, and *latency* and *pointing to the target object* was collected during the test phase. At visit 1, children were tested on their *in-depth learning* of the target word immediately after the video using measures of *word association*, *word description*, and *word generalization*, in that order. At visit 2, children were only tested on their in-depth learning and did not watch the video again. *Word production* was assessed at the end of each visit.

Analysis Plan

This study investigated how children with ASD and TD children use a gaze versus an arrow cue. Therefore, results focused on the matched group of ASD and TD children in a 2 (cue condition: gaze versus arrow) x 2 (group: ASD vs TD) mixed design. An additional factor was included when measures occurred over multiple time points (i.e., video phases, visits).

Continuous measures such as eye tracking variables, the number of semantic features, and word generalization accuracy and response times were evaluated using linear mixed effects models with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015b) in R. Mixed effects models were used to analyze the data because: 1) the present data violated assumptions of sphericity required in a classical mixed analysis of variance (ANOVA), and 2) mixed models can accommodate unbalanced grouping variables due to any dropped children (e.g., children who were unable to calibrate for eye tracking data) and missing data (e.g., dropped video phases).

Mixed models offer many benefits over the classical ANOVA, such as accounting for random effects (e.g., participant, object) and fixed effects, (e.g., condition, group) over repeated observations. Additionally, mixed models can include covariates, which in this dissertation refers to continuous variables that are not considered to be fixed effects of interest (e.g., IQ, age). However, in this dataset each cue condition included only 2 observations per individual, and there were 24 children in each group, which limits exploring all random effects, fixed effects, and covariates. Therefore, covariates such as IQ and age were excluded since groups were already well matched. Following the recommendations by Bates, Kleigl, Vasishth, and Baayen (2015a), I used the most parsimonious models regarding random effects since the main focus was on factorial contrasts, which were fixed effects of cue condition and group. Parsimonious models included only a random intercept of participant. Justification for excluding additional random effects (e.g., for objects and all potential slopes), and model building can be seen in Appendix F.

For the final models, assumptions of normally distributed errors (difference between observed data and data predicted by the model) for random effects and the model, and homoscedasticity were satisfied by visual inspection of quantile-quantile (q-q) plots and histograms. If homoscedasticity appeared to be violated, the dependent variable was transformed to attempt to improve the violation. Square root transformations were used with positively skewed data that included many zero values. Log transformations were used for positively skewed data when there were values greater than zero, because log transformations are known to better correct for skewness than square root transformations. Power transformations were used for negatively skewed data (i.e., when there was a ceiling effect). P values for mixed models were calculated using the lmerTest package with the Kenward-Roger method to calculate degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2016).

Post hoc tests were conducted on effects when $p < .05$, and were calculated with the lsmeans package (Lenth, 2016). The *p* values of multiple comparisons were corrected with the Tukey method and a family-wise error of alpha = .05 (Lenth, 2016). The lsmeans package provides predicted marginal means (*PMM*s), which are means adjusted for when data are unbalanced (e.g., missing data). For all transformed data, *PMM*s were back-transformed in lsmeans and thus can be interpreted in their original units. The difference scores of *PMM*s are similar to other difference scores between means and can be considered as unstandardized effect sizes. Standardized effect sizes are provided by Cohen's *d* (Westfall, 2016) based on the means and standard deviations of the raw data (data not corrected for unbalanced data), not the *PMM*s and *SE*s used by the model. All raw data are provided in tables with each research question for transparency. Cohen's *d* was calculated with the mes2 function of the compute.es package (Del Re, 2013) using a pooled standard deviation for the denominator for between and within-subjects variables (Lakens, 2013; Westfall, 2016).

Potential outliers were examined by calculating Cook's distance with the influence.ME package for mixed models (Nieuwenhuis, Grotenhuis, & Pelzer, 2012). Outliers were numerically identified as participants with Cook's distance values greater than 4/*n*, where *n* in our case refers to the number of participants (Belsley, Kuh, & Welsch, 1980). Cook's distance provides a summary score for each participant's influence on all of the parameters specified in the model. Cook's distance lets us determine the influence of that participant on the variables included in the model. Identifying influential points in mixed effect models are important, especially when there may be few observations (Nieuwenhuis et al. 2012). Influential points take into consideration both extreme residuals (difference between the predicted and actual observation), and also extreme scores on the independent variables (e.g., extreme scores with respect to the group of children or cue condition). It could be possible that a participant provided scores that are considered extreme scores on the independent variable, but is fitted well by the model, thus would not be considered an influential point. There are multiple ways of dealing with influential points (when these points are not due to errors in the data), including excluding the points, obtaining additional data points to account for overly influential ones, adding additional variables to the model to improve model fit. In this dissertation, I checked if these influential points were due to errors in the data, but if there were no errors that could be found then I excluded these points. Given that influential data points could affect potential inferences that could be made from the model (Nieuwenhuis et al., 2012), and influential points are important to consider when there are fewer observations, I presented data both with and without outlier participants. For each research question below, when there were outlier participants, analyses were also run without outliers. For most models, results of significant and non-significant effects were similar with and without outliers, and in these cases all participants were kept in the final model. For measures of latency and word description, results differed with and without outliers, thus both models were presented in the research questions below.

When data were categorical or had a limited range (i.e., 2 possible correct for pointing in each cue condition), each child was placed into a category that summarized his/her *level of success at learning*. The categories were summarized to highlight whether children were or were not able to learn and if there were differences in whether children learned differently with gaze or an arrow cue, which was one of the main factors of interest in this study. Levels of success at learning were summarized into five levels which can be seen in Table 5. These levels included: 1) *success with both cues,* meaning children identified the target in three or four trials, 2) *success with gaze* or 3) *success with arrow*, meaning children identified the target in only both trials of the respective cue condition, 4) *at chance with both cues,* meaning children only identified the target once in the gaze and once in the arrow condition, and 5) *limited or no success*, meaning children could only identify the target in one or none of the trials. The distributions between groups of children were evaluated using Fisher's exact test.

Table 5. Levels of Success at Learning

Gaze condition (number correct)	Arrow condition (number correct)	Level
2	2	Success with both cues
2	1	
1	2	
2	0	Success with gaze
0	2	Success with arrow
1	1	At chance with both cues
1	0	Limited or no success
0	1	
0	0	

Given 3 levels of response (2, 1, or 0 correct) for each of 2 cue (gaze, arrow), there were 3 x 3 = 9 possible response categories. Nine possible response categories were grouped into the 5 learning levels shown above.

Eye Tracking Diagnostics

Checking eye tracking data is necessary to validate the precision and accuracy of the data. Two main steps to validate the data include verifying calibration and data cleaning to account for data when children were not looking long at the video, or low looking times, which are defined differently across studies. Calibration occurs during testing, whereas data cleaning occurs after the data has been collected. A short summary is provided below, but please refer to Appendix G for a more detailed report.

Calibration. FaceLAB provides a mean angular error score for each eye after calibration, which is a value that represents the accuracy and precision of the movements of their measurement. Using a visual angle calculation, which includes the child's distance from the screen, a mean angular error score of 5 was determined as an acceptable cut-off for the current study. This mean angular error score ensured that all children were held to the same standard of accuracy and precision that least compromised the integrity of the data for the main research questions (e.g., data points for attention to the gaze area could not be mistaken for data points for attention to the arrow area). If children did not meet this criterion, they continued to watch the videos but their eye tracking data were not analyzed. Children were calibrated with a 9-point calibration process and within 5 degrees mean angular error for each eye; all children had to meet the mean angular error for both eyes. Groups of children did not significantly differ on their mean angular error, suggesting that the accuracy and precision of eye tracking data were similar between groups. Two children with ASD were not included in eye tracking analyses because 1 child had strabismus and the other could not be calibrated; this resulted in 22 children with ASD and 24 TD children included in eye tracking analyses.

Data cleaning. One important part of data cleaning entails decisions about how to account for overall low looking times to videos (Venker & Kover, 2015). Overall low looking times, assuming calibration criteria were met, can be due to excessive movement or other technical issues with the eye tracker itself, as well as low looking by the child. Looking time information was available for each video phase, which is a value summarized from smaller units of looking time. These smaller units can be defined in different ways. For all research questions, I adopted a unit of 100 ms as the smallest unit of looking time, which is commonly used as an indicator of children's cognitive processing (Gredebäck, Johnson, & Hofsten, 2009; Oakes, 2012), and will hereby be referred to as a *fixation*. However, for overall looking time, I used a unit of 33 ms because I wanted to use a more fine-grained measure for children's overall looking time to each phase.

An overall looking time cut-off of 25% looking relative to the duration of the phase was adopted for eye tracking data in the current study. One video sequence included baseline, teaching (label and cue portions), center, and test phases, and each child had 4 video sequences (2 per cue condition). Video sequences were considered complete for analyses when overall looking times were above 25% to baseline, cue portion of the teaching phase, and test phases. Video sequences were dropped when they were not considered complete for analyses. Four video sequences were dropped for children with ASD (affected 4 different children) and 6 video sequences were dropped for TD children (affected 4 different children) because they did not meet criteria for complete video sequences. One TD child did not have full video sequences for both cue conditions and was removed from eye tracking analyses. Therefore, after verifying eye tracking diagnostics, the sample included in eye tracking analyses included 22 children with ASD and 23 TD children from the matched group.

I used a linear mixed model to examine children's overall looking time to the scene using the dependent variable of the proportion of overall looking time. The proportion was calculated by dividing the overall looking time in a phase (based on a unit of 33 ms) relative to the duration of that phase; proportions ranged from 0 to 1. There were no significant differences between groups of children or cue conditions. Children looked for a significantly smaller proportion of overall looking time during test in comparison to the baseline and teaching phases, which indicated that children's overall attention to the scene decreased across video phases. However, the mean proportion of overall looking time of 78% during the test phase, still seemed like children were looking during the majority of the test phase (raw $M = .78$, $SD = .20$, range = 25.21 – 100), despite relatively less looking in comparison to the baseline phase. Please refer to Appendix G for more details regarding eye tracking diagnostics.

RESULTS

All visual attention measures below were calculated using a unit of 100 ms fixation duration.

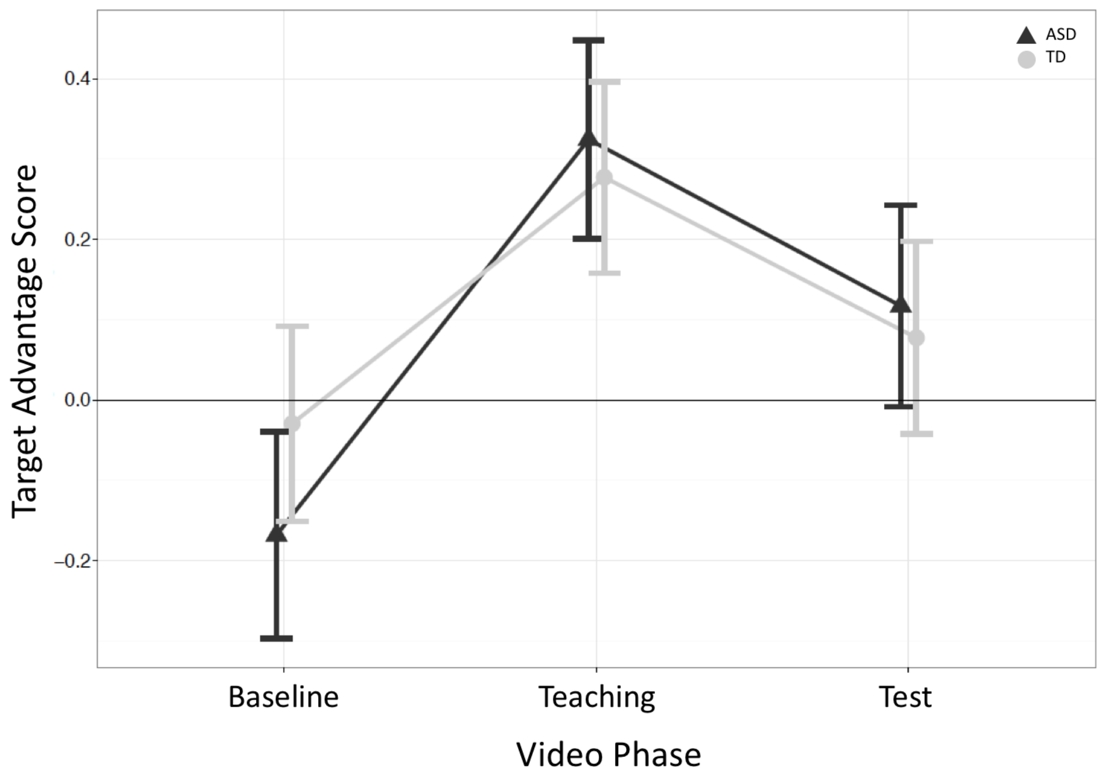
Visual Attention to the Target Object: Target Advantage

Children's ability to follow the cue to the target object during teaching (cue portion only) and locate the target during test was measured using children's fixation duration to the target relative to the distractor. This was calculated using a standardized difference score to create a *target advantage score* (Akechi et al., 2011; Senju, Southgate, White, & Frith, 2009) during baseline, teaching, and test phases. First, the total fixation duration to the target and distractor objects were summarized in each baseline, teaching and test phase. Then, target advantage scores were calculated by subtracting the fixation duration to the distractor (d) from the fixation duration to the target (t), and dividing by the total fixation duration to both the target and distractor, i.e, (t – d) / (t + d). Scores ranged from -1 to 1, with positive numbers indicating more looking to the target and negative numbers indicating more looking to the distractor. Scores were calculated for baseline, teaching, and test phases, and provide a way to monitor attention to the target across the changing phases of the video. That is, higher scores during teaching versus baseline could indicate that children followed the cue during teaching and higher scores during test in comparison to baseline indicate that children may have been influenced by the cue during teaching, thus locating the target during test. In addition to using this measure to monitor children's attention to the target, our primary analysis examined whether looking time to the target differed between cue conditions, and if there were any interactions of this factor with group or video phase.

The linear mixed model included a random intercept of participant and fixed effects of cue condition (gaze, arrow), group (ASD, TD), and video phase (baseline, teaching, test). There were no significant main effects of cue condition, group, or interactions between cue, group or video phase (Fs < 1.81, ps > .17). However, there was a significant effect of video phase, $F(2, 448.83) = 23.46$, $p < .001$. Post hoc tests revealed that children demonstrated more of a target advantage during teaching (*PMM* = .30, *SE* = .04) than they did during baseline (*PMM* = -.10, *SE* = .04), $t(469.64) = -6.85$, $p < .001$, *PMM* difference = - .40, $d = -.66$. Moreover, children had more of a target advantage during test (*PMM* = .10, *SE* = .04) versus baseline, $t(468.99) = -3.34$, $p = .003$, *PMM* difference = - .20, $d = -.30$. These findings suggest that children in both groups were influenced by the cues during teaching. Children also had less of a target advantage during test than they did during teaching, $t(468.73) = 3.53$, $p = .001$, *PMM* difference = .20, $d = .33$, which was expected since there was no cue during test to direct their attention.

Thus, while watching the baseline, teaching and test phases, children with ASD and TD children used both gaze and arrow cues to look at the target during teaching and correctly visually locate the target object during test. Negative target advantage scores during baseline indicated an initial preference for the distractor. This pattern suggests that following the cue during teaching required many children to break their focus of attention from the distractor to follow the cue to the target object. Raw target advantage scores by cue condition, group, and video phase can be seen in Table 6. Figure 7 below collapses across cue condition to depict looking pattern over time in each group of children.

Figure 7. Word Learning: Predicted Marginal Means of Target Advantage Scores Collapsed Across Cue Conditions for Each Video Phase in Each Group of Children



Points represent predicted marginal means and error bars are 95% confidence intervals. More looking to the target is shown with positive values and more looking to the distractor is shown with negative values. The horizontal line of 0 indicates equal looking to the target and the distractor.

Table 6. Word Learning: Raw Target Advantage Scores

Baseline	ASD (n = 22)	TD (n = 23)	d
Gaze	-.20 (.62)	.08 (.61)	-.44
Arrow	-.14 (.48)	-.14 (.58)	-.01
d	-.09	.36	
Teaching			
Gaze	.30 (.52)	.20 (.52)	.20
Arrow	.35 (.56)	.36 (.43)	-.03
d	-.09	-.35	
Test			
Gaze	.07 (.56)	.01 (.64)	.11
Arrow	.16 (.56)	.15 (.57)	.02
d	-.15	-.23	

The values shown are the mean(*SD*). Ranges were – 1 to 1 for all subgroups except for baseline ASD arrow condition (-1 to .73), teaching TD arrow condition (-.72 to 1) and test ASD gaze condition (-.84 to 1). Negative Cohen's *d* values represent higher scores for typically developing children when comparing across groups and higher scores for arrow when comparing across cue conditions.

Visual Attention to the Cue: Proportion of Looking Time and Contingent Looking

Children's attention to the cue area was measured with the *proportion of looking time to the cue area* and the number of looks between the cue area and the target object, or contingent looks (Norbury et al., 2010; Tenenbaum et al., 2014).

Proportion of looking time to the cue area. First, the total fixation duration to the cue area and the total fixation duration to the scene was summarized in baseline and teaching phases. The proportion of looking time to the cue area was calculated by dividing the total fixation duration to the cue area by the total fixation duration to the scene, resulting in proportions ranging from 0 to 1. Both baseline and teaching phases were included for this measure because of the relevance of direct gaze (present in the baseline phase of the gaze condition, as seen in

Figure 3). Direct gaze is a salient ostensive cue (Csibra & Gergely, 2009), and responses to direct gaze have been found to differ between children with ASD and TD children (Chawarska, Macari, & Shic, 2012; Norbury et al., 2010). Though comparing direct and referential gaze is not a goal of the current study, it may be important to the interpretation of potential condition effects. For example, any differences between groups during teaching could be due to a similar pattern of pre-existing differences in their initial attention at baseline (to the actor's direct gaze). It should be noted that in the arrow condition, the control for direct gaze was a static circle during baseline. Please refer to Figure 2 in the Methods section for a visual representation of this contrast. For the purpose of this analysis I will refer to the area of direct and referential gaze as the area of the gaze cue, and the area of the static circle and arrow as the area of the control cue.

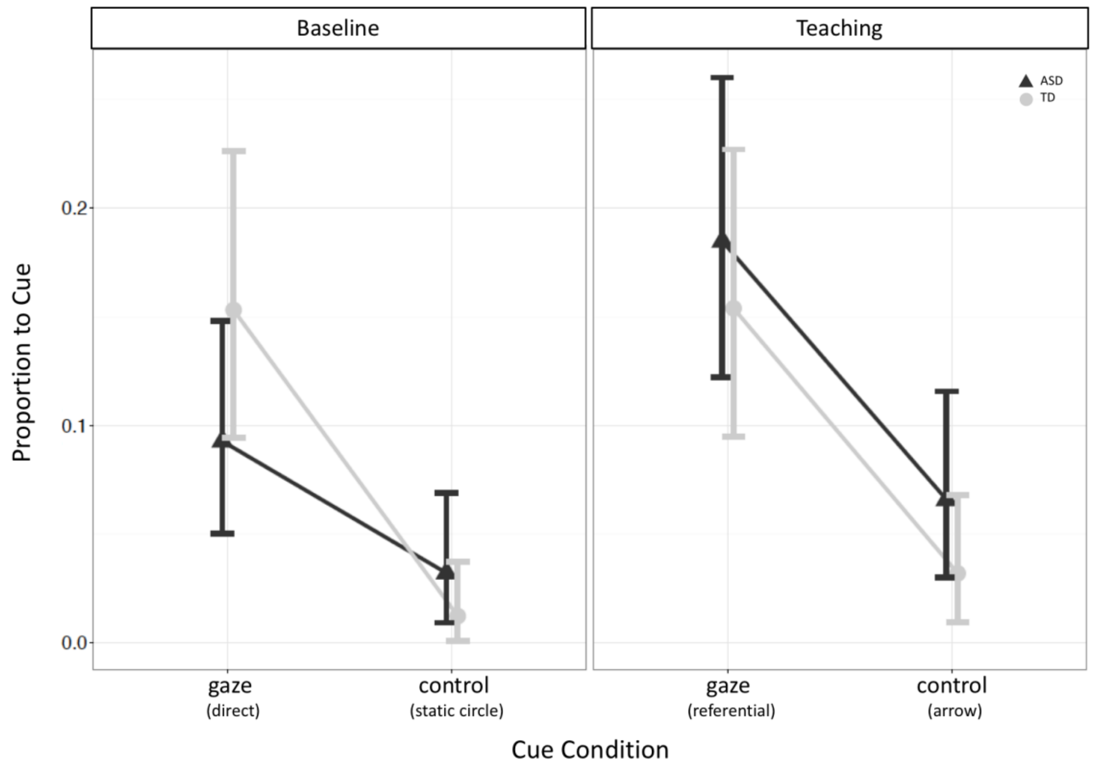
The linear mixed model included a random intercept of participant and fixed effects of cue condition (gaze, arrow), group (ASD, TD), and video phase (baseline, teaching). There was heteroscedasticity due to children with 0% time looking to the cue area. To try and correct for the heteroscedasticity, I first removed data from 4 children (3 TD, 1 ASD) who never looked at the cue area (though they met data cleaning criteria of more than 25% overall looking to the scene), and then performed a square root transformation on the proportion of looking time to the cue area. As seen in Table 7, there was a significant main effect of video phase, with all children spending a significantly greater proportion of looking time to the cue area during teaching (*PMM* = .10, *SE* = .02) than during baseline (*PMM* = .06, *SE* = .01), *PMM* difference = - .04, *d* = -.17. There was also significant main effect of cue condition, and a 2-way interaction between cue condition and group. A post-hoc test of the 2-way interaction found no differences between groups of children (*p*s > .46), and confirmed the cue condition main effect within each group of children. Children with ASD spent a significantly greater proportion of looking time to the area of the gaze cue versus the control cue (gaze *PMM* = .13, *SE* = .03; control *PMM* = .05, *SE* = .02), $t(280.03) = 4.64$, $p < .001$, *PMM* difference = .08, $d = .46$. The same pattern was seen in TD children (gaze *PMM* = .15, *SE* = .03; arrow *PMM* = .02, *SE* = .01), $t(279.10) = 7.60$, $p < .001$, *PMM* difference = .13, $d = .85$. Figure 8 depicts the cue condition effect and video phase effect in each group of children. Raw proportion of looking time to the cue area by cue condition, group, and video phase can be seen in Table 8.

Table 7. Word Learning: Main Effects and Interactions of Proportion of Looking Time to Cue Area

Factor	*df*	*F*	*p*
Cue condition (C)	1, 267.26	74.92	< .001***
Group (G)	1, 39.00	.31	.58
Video phase (V)	1, 265.04	8.91	.003**
C x G	1, 267.26	4.56	.03*
C x V	1, 265.04	.05	.83
G x V	1, 265.04	2.16	.14
C x G x V	1, 265.04	1.59	.21

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

Figure 8. Word Learning: Predicted Marginal Means of Proportion of Looking Time to Cue Area for Each Video Phase and in Each Group of Children



Points represent predicted marginal means and error bars are 95% confidence intervals. To ease interpretation, the proportion to cue variable was back transformed, thus units represent proportion scores and not the square root transformation.

Table 8. Word Learning: Raw Proportion of Looking Time to Cue Area

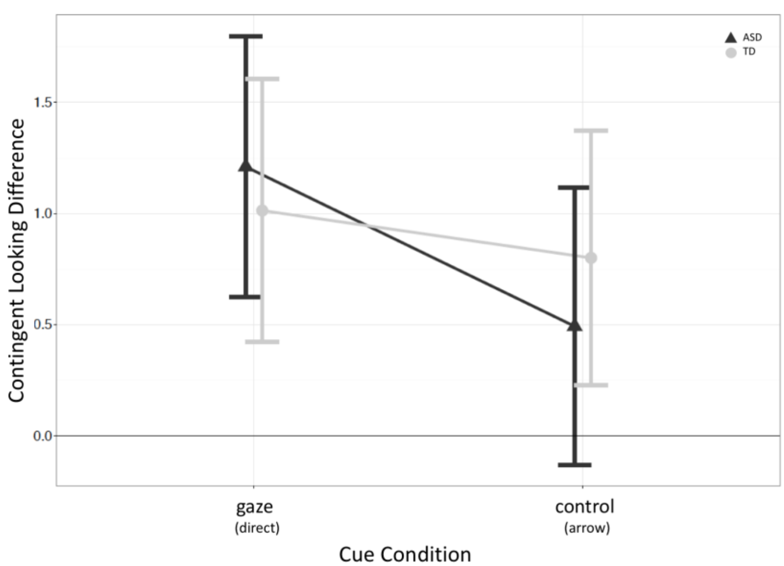
Baseline	ASD (n = 21)	TD (n = 20)	*d*
Gaze	.16 (.18)	.21 (.19)	-.29
	0 - .71	0 - .80	
Arrow	.09 (.15)	.05 (.12)	.28
	0 - .77	0 - .48	
d	.42	1.01	
Teaching			
Gaze	.24 (.17)	.22 (.18)	.11
	0 - .68	0 - .81	
Arrow	.12 (.14)	.07 (.10)	.46
	0 - .63	0 - .35	
d	.75	1.07	

The values shown are the mean(*SD*) and range. Negative Cohen's *d* values represent higher scores for typically developing children when comparing across groups and higher scores for arrow when comparing across cue conditions.

Contingent looking. First, I measured the shifts between looking at the cue area and the target or distractor object (Baldwin, 1993b; Norbury et al., 2010). I calculated contingent looking with fixations, that is, one contingent look was defined by a shift from a fixation on the cue area followed immediately by a fixation on the target or distractor object area, and vice versa. For example, if the child looked from the cue area to the target, then back to the cue area, this counted as two contingent looks. Next, the total number of contingent looks between the cue area and distractor was subtracted from the total number of contingent looks between the cue area and the target. Therefore, this difference score was called a *contingent looking difference* score. A standardized difference score was not calculated because there were many instances of a difference score of 0, which was either because children did not provide any contingent looks to neither the target nor the distractor, or there were equal contingent looks to the target and distractor. A standardized difference score would result in dividing 0 over 0, which is not a real number. All 0 values were retained except for 6 children (3 ASD, 3 TD) who had no contingent looks for all video sequences. Data were only analyzed during the teaching phase, when the gaze or arrow cue directed attention to the target.

The linear mixed model included a random intercept of participant and fixed effects of cue condition (gaze, arrow) and group (ASD, TD). There was a marginal, though non-significant effect of cue condition, $F(1, 109.17) = 3.10$, $p = .08$. The main effect of group, $F(1, 36.92) = .03$, $p = .87$, and the interaction between cue condition and group was not significant, $F(1, 109.17) = .90$, $p = .34$. Thus although for the proportion of looking to cue area there was more looking to the gaze versus the control area, there did not appear to be a comparable significant effect with contingent looking. Figure 9 depicts the contingent looking difference in each group of children. The raw data by group and cue condition can be seen in Table 9.

Figure 9. Word Learning: Predicted Marginal Means of the Contingent Looking Difference in Each Group of Children



Points represent predicted marginal means and error bars are 95% confidence intervals. More contingent looking to the target is shown with positive values and more contingent looking to the distractor is shown with negative values. The horizontal line of 0 indicates an equal number of contingent looking to the target and the distractor.

Table 9. Word Learning: Raw Contingent Looking Difference Scores

Teaching	ASD (*n* = 19)	TD (*n* = 20)	*d*
Gaze	1.21 (2.09)	1.00 (2.23)	.10
	-3 to 6	-5 to 5	
Arrow	.48 (1.73)	.80 (1.69)	-.18
	-3 to 4	-1 to 6	
d	.38	.10	

The values shown are the mean(*SD*) and range. Negative Cohen's *d* values represent higher scores for typically developing children when comparing across groups and higher scores for arrow when comparing across cue conditions.

These visual attention measures to the cue area demonstrate no significant differences between groups of children on either measure. For both groups of children, while there is a significant difference between cue conditions in children's proportion of looking time to the cue area, with more looking at the cue area in the gaze relative to the arrow condition, there is no significant cue condition difference on their contingent looking difference. For the proportion of looking time to the cue area, children also spent more time looking at the cue area during the teaching phase, when the cue indicated an object, relative to the baseline phase.

Visual Attention to the Target Object During Test: Latency

Latency, or how quickly children first fixated (i.e., first time children spend a 100 ms on the target object) to the target object at test, was measured from the start of the test video, which coincided with the end of the prompt (e.g., "*Where is the pagoune*?") during the ISI[4]. Latencies

[4] In contrast to other methods of calculating latencies, in the current study, latencies were included for children's first fixation on the target object, regardless of whether children fixated on the distractor prior to the target object (Aldaqre et al., 2015; Bani Hani et al., 2012; Fernald, Perfors, & Marchman, 2006; Hurtado, Marchman, & Fernald, 2008). The inclusion of all first

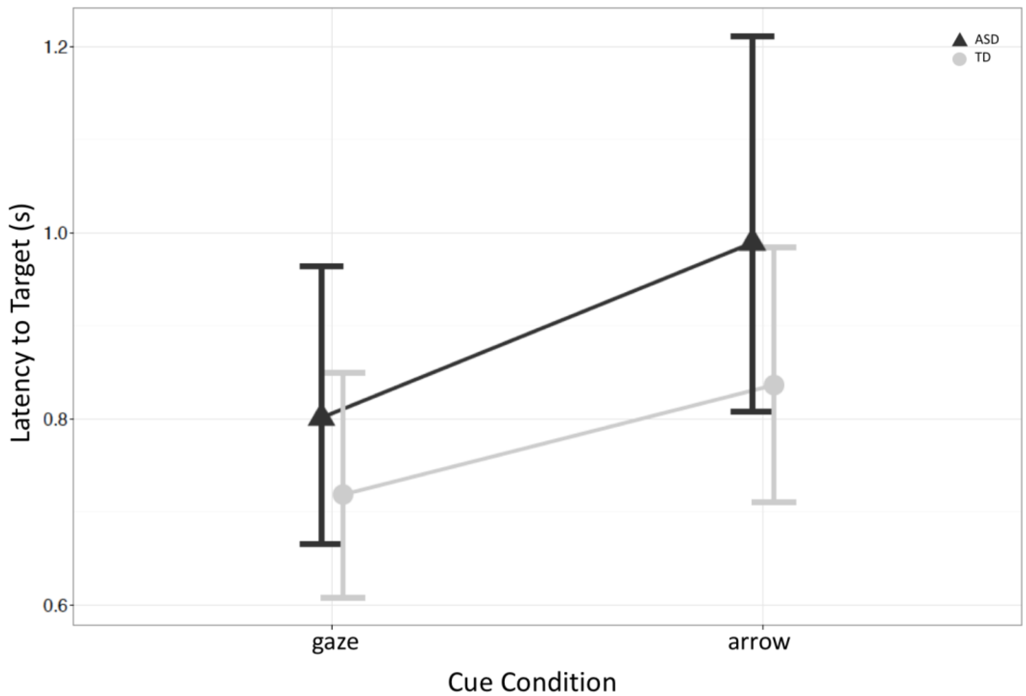
under 200 ms were excluded based on the duration it takes to make a saccade (Altmann, 2011), thus the time needed to switch attention to the target rather than accidentally looking at it. Latencies over 4000 ms were excluded from analyses because at 4000 ms children heard the beginning of the second prompt to point to the target, e.g., "*Now point to the pagoune*." These exclusion criteria resulted in removing all trials from 1 TD child.

The linear mixed model included a random intercept of participant and fixed effects of cue condition (gaze, arrow) and group (ASD, TD). Latencies were log transformed due to heteroscedasticity. Results differed with and without outlier participants, determined using Cook's distance. Including all children, there were no significant main effects of cue condition, group, or interaction between cue condition and group (Fs < 1.94, ps > .17). Removing outliers (4 ASD) resulted in a significant effect of cue condition, $F(1, 106.88) = 4.02$, $p = .047$, with faster latencies to the target in the gaze (*PMM* = .76, SE = .05) versus the arrow condition (*PMM* = .91, SE = .06), *PMM* difference = -.15, $d = -.15$. This cue condition difference is depicted in Figure 10, and raw latencies can be seen in Table 10 by cue condition and group. There were no significant group or interaction effects between cue condition and group (Fs < 2.30, ps > .14). However, it should be noted that the model without outliers reflects 18 children with ASD and 22 TD children, which is smaller than our original sample of 22 children with ASD and 23 TD children. Nevertheless, these findings demonstrate cue condition differences, with faster fixations to the target in the gaze relative to the arrow condition in both children with ASD and typically-developing children.

---

fixations was because the prompt in this study was provided prior to children seeing both objects versus while both objects were presented to children.

Figure 10. Word Learning: Predicted Marginal Means of Latency to the Target Object During Test in Each Group of Children – Removing Outliers



Points represent predicted marginal means and error bars are 95% confidence intervals. To ease interpretation, the latency to target variable was back transformed, thus units represent ms and not the log transformation.

Table 10. Word Learning: Raw Latency Scores to the Target During Test (ms) – Removing Outliers

	ASD (*n* = 18)	TD (*n* = 22)	*d*
Gaze	.99 (.85)	.82 (.62)	.20
	.23 – 3.25	.36 – 3.67	
Arrow	1.09 (.52)	.93 (.55)	.29
	.45 – 2.65	.30– 2.29	
d	-.13	-.19	

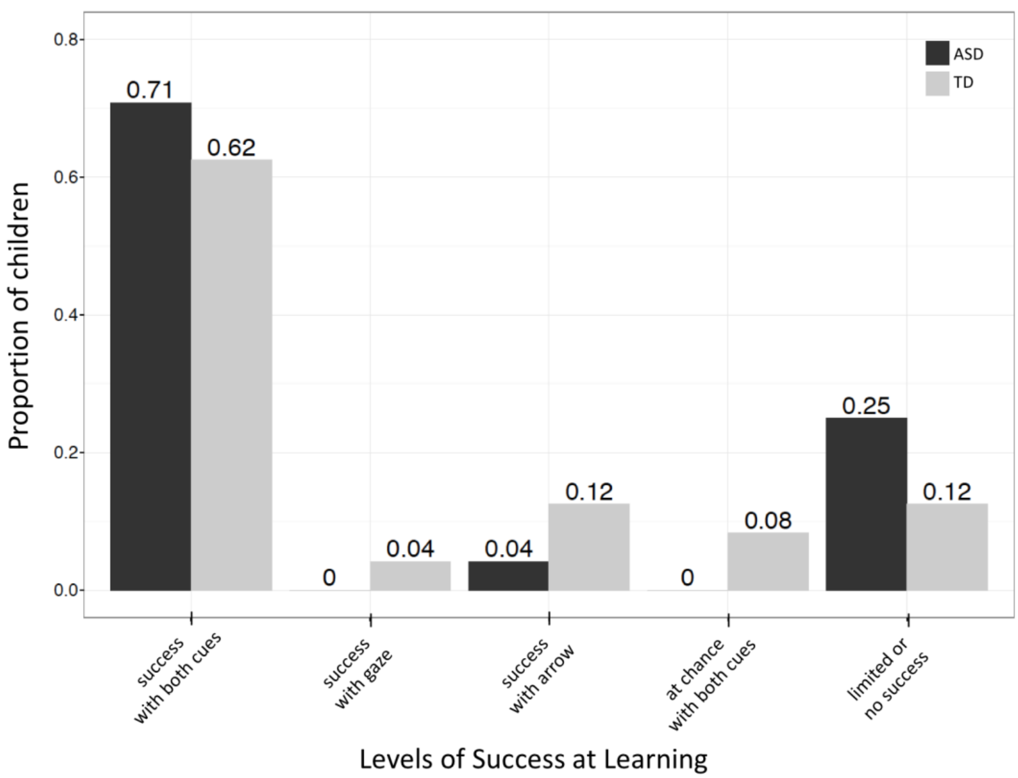
The values shown are the mean(*SD*) and range. Negative Cohen's *d* values represent higher scores (slower) for typically developing children when comparing across groups and higher scores (slower) for arrow when comparing across cue conditions.

Learning: Word Recognition

At the end of the test phase, children were asked to point to the target object as a measure of explicit word recognition. Success at learning was summarized into five levels noted earlier in Table 5: *success with both cues*, *success with gaze*, *success with arrow*, *at chance with both cues*, or *limited or no success*. All children pointed correctly in practice trials to familiar words, demonstrating that they understood the procedure, thus all 24 children with ASD and 24 TD children were included in this measure.

There were no significant differences between groups, Fisher's exact test $p = .31$. The majority of children in both groups had success with both cues, meaning that they pointed to the target correctly for three or four trials (17 ASD, 15 TD). Very few children had success only with gaze (0 ASD, 1 TD) or an arrow cue (1 ASD, 3 TD). Remaining children were at chance with both cues (0 ASD, 2 TD), or had limited or no success with either cue (6 ASD, 3 TD). The distribution of levels of success at learning in each group of children can be seen in Figure 11, which depicts the proportion of children in each group for ease of comparison with other measures. These findings indicate that there was no difference between groups in the distribution of their levels of success for their explicit word recognition, and neither group included children who only had success with gaze or an arrow cue.

Figure 11. Word Learning: Children's Levels of Success at Learning for Pointing



This figure depicts the proportion of children that were in each level of success at learning on the measure of their pointing during the test phase.

In-depth Learning Immediately After the Video and One Week Later

Word association. After pointing, children were asked to provide word associations to the target word and familiar words. This measure was also conducted one week later. Word association was included as an exploratory measure to examine children's semantic knowledge (i.e., what the object looks like, how it can be used) of the target words. There were six possible coding groups: semantic, clang, distractor, any experimental object, error and other. As seen in prior studies that have coded word associations (Sheng & McGregor, 2010), for the purpose of this analysis, children's learning was defined only using semantic responses, whereas responses in all other coding groups were not considered as providing semantic knowledge of the target word. Semantic responses were the most frequent response at visit 1, though this decreased at visit 2. At each visit, I counted the number of semantic responses that children provided per cue condition and summarized responses into the five levels of success at learning. Before providing results from target words, I first provide results of semantic responses from children's familiar words, to examine if there are children who had difficulty with the procedure and thus should be removed from further analyses of the word association task.

Familiar words. An evaluation of children's responses to the familiar words (10 familiar words at each visit) indicated that both groups of children seemed to understand the task, because many children provided a minimum of 7 semantic responses to familiar words at both visits (18 ASD, 21 TD). Seven children with ASD provided a minimum of 2 semantic responses to familiar words and 3 TD children provided a minimum of 3 semantic responses. Only 2 children with ASD did not provide any semantic responses for familiar words, but did provide clangs (alliteration or shared a similar sound). Because the use of clangs suggests phonological associations, which are related to the sound of the word rather than demonstrating a complete lack of responding in the task, these children were included for analyses of their responses to target words. For 1 child with ASD, the task was discontinued during the first visit due to the child's limited expressive language ability. Therefore, the following analyses were based on 23 children with ASD and 24 TD children.

Target words. For semantic responses to target words, there were no significant differences between groups of children, Fisher's exact test $p = .13$, or at visit 2, Fisher's exact test $p = .55$. At visit 1, there were some children in both groups that had success with both cues, meaning that semantic responses made up three or all four trials (8 ASD, 5 TD). However, almost half of the children in both groups had limited or no success, with only one or no semantic responses (11 ASD, 14 TD). As seen above with word recognition, there were children who had success with only gaze (1 ASD, 2 TD) or the arrow (3 ASD, 0 TD). Three TD children were at chance with both cues, meaning that they provided one semantic response in each cue condition. Children provided fewer semantic responses over time. By visit 2, most children had limited or no success (17 ASD, 20 TD). Only 1 child with ASD, and no TD children had success with both cues. Again, few children had success only with gaze (1 ASD, 2 TD), the arrow (none) or were at chance with both cues (4 ASD, 2 TD). These findings of children's word associations demonstrate that there were no differences between groups in their one-word semantic responses to the target word, and neither group included children who had success with only gaze or the arrow cue. However, in contrast to word recognition, there were much fewer children who had success with both cues.

When children did not provide semantic responses, their responses varied widely. Appendix H provides a complete depiction of individual children's responses at each visit and descriptive data can be seen in Table 11. At visit 1, the fewest responses were in the distractor coding group, whereas coding groups of clang, any experimental object, other, and error were fairly similar in the number of responses. At visit 2, coding groups of semantic, clang, and error were similar in the number of responses, with distractor again including the fewest responses. Clang responses were similar in both groups of children. Immediately after the video, 2 children with ASD provided three or four clang responses and one week later 3 children with ASD and 2 TD children provided three or four clang responses. These findings suggest that at visit 1 and 2, no other coding categories were favored more than another when children were not providing semantic responses.

Table 11. Word Associations: Number of Associations in Each Category

Visit 1	ASD (n = 23)	TD (n = 24)
Semantic	1.70 (1.46)	1.42 (1.18)
	0 – 4	0 – 4
Clang	.57 (1.12)	.42 (65)
	0 – 4	0 – 2
Distractor	.04 (.21)	.13 (.45)
	0 – 1	0 – 2
Any experimental object	.48 (.95)	.67 (1.01)
	0 – 3	0 – 3
Other	.74 (.86)	1.04 (.91)
	0 – 3	0 – 3
Error	.43 (.99)	.33 (.87)
	0 – 4	0 – 4
Visit 2		
Semantic	0.96 (.88)	0.75 (.74)
	0 - 3	0 - 2
Clang	.87 (1.25)	.79 (1.14)
	0 – 4	0 – 4
Distractor	.04 (.21)	.08 (.28)
	0 – 1	0 – 1
Any experimental object	.61 (.89)	.54 (1.06)
	0 – 3	0 – 4
Other	.48 (.67)	1.21 (1.22)
	0 – 2	0 – 4
Error	1.0 (1.35)	.63 (1.06)
	0 – 4	0 – 3

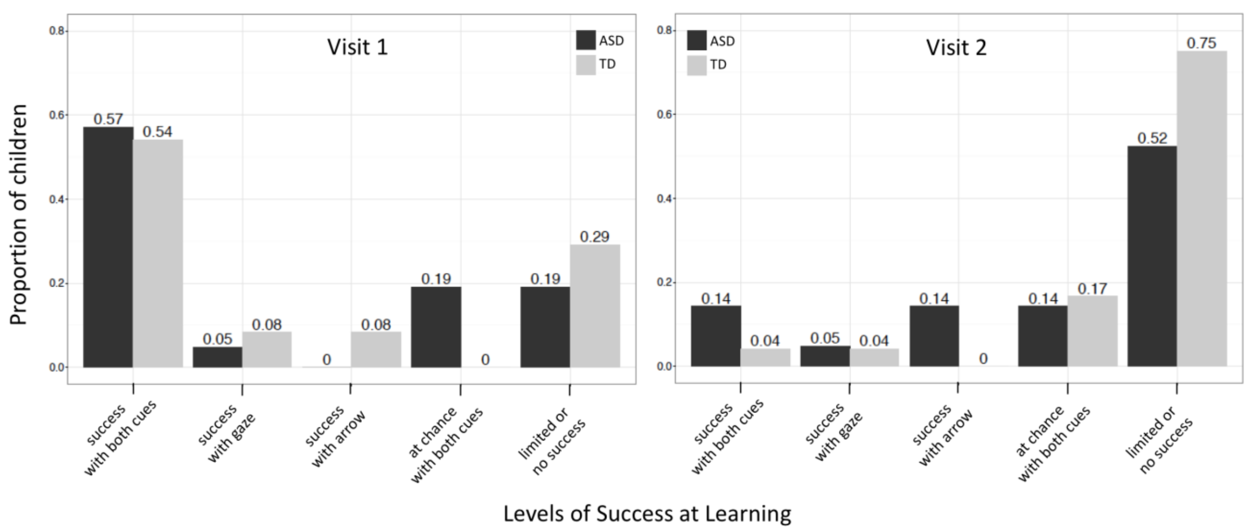
The values shown are the mean(*SD*) and range.

Word description. After word association, children were asked to describe the target object. Like word association, word description was included to examine children's semantic knowledge (i.e., what the object looks like, how it can be used), but in more detail than one-word responses by using the number of semantic features recalled when describing the object. First, I examined if there were descriptions of familiar objects that were unable to be identified by the coder. If children could not provide valid descriptions of familiar objects, then children were excluded from this analysis because this indicated that children had difficulty with the procedure and may not have understood the task. Next, between groups of children, I compared the distribution of levels of success at learning for the number of valid target object descriptions (see Methods for description of what counted as a valid description). Finally, I examined the number of semantic features when children provided valid target object descriptions.

Familiar words. At each visit, children provided descriptions for two familiar objects. Children were only included when they provided valid descriptions for both familiar objects at both visits. These criteria excluded 3 children (3 ASD), therefore, 21 children with ASD and 24 TD children were included in analyses with target words.

Target words – number of valid descriptions. There were no significant differences between groups in their levels of success at learning at visit 1, Fisher's exact test $p = .13$, or visit 2, Fisher's exact test $p = .23$. Figure 12 depicts the proportion of children in each level. About half of the children in both groups had success with both cues, describing the target object in three or four trials (12 ASD, 13 TD). When children did not have success with both cues, there were few who had success with only gaze (1 ASD, 2 TD), or the arrow cue (0 ASD, 2 TD). There were more children with ASD who were at chance with both cues, describing the target once in the gaze condition and once in the arrow condition (4 ASD, 0 TD), and a similar number of children in both groups who had limited or no success with either cue (4 ASD, 7 TD). As seen in word association, children had more difficulty by the second visit. At the second visit, many of children's descriptions were unable to be identified as valid, demonstrating that, one week later, it may have been more difficult to recall semantic features: success with both cues (3 ASD, 1 TD), success with only gaze (1 ASD, 1 TD), success with only the arrow (3 ASD, 0 TD), at chance with both cues (3 ASD, 4 ASD), limited or no success with either cue (11 ASD, 18 TD). These findings exhibit that in both visits there were no cue condition differences, with more children at visit 1 having success with both cues or limited or no success. However, by visit 2 there were more children having limited or no success, which means that findings regarding semantic features include more data per child from visit 1 than visit 2.

Figure 12. Word Description: Children's Level of Success at Learning for Valid Target Object Descriptions



This figure depicts the proportion of children that were in each level of success at learning on the measure of their valid target object descriptions, part of the word description task. Valid target object descriptions were determined by a coder (see Methods and Appendix for more detail on what counted as a valid description).

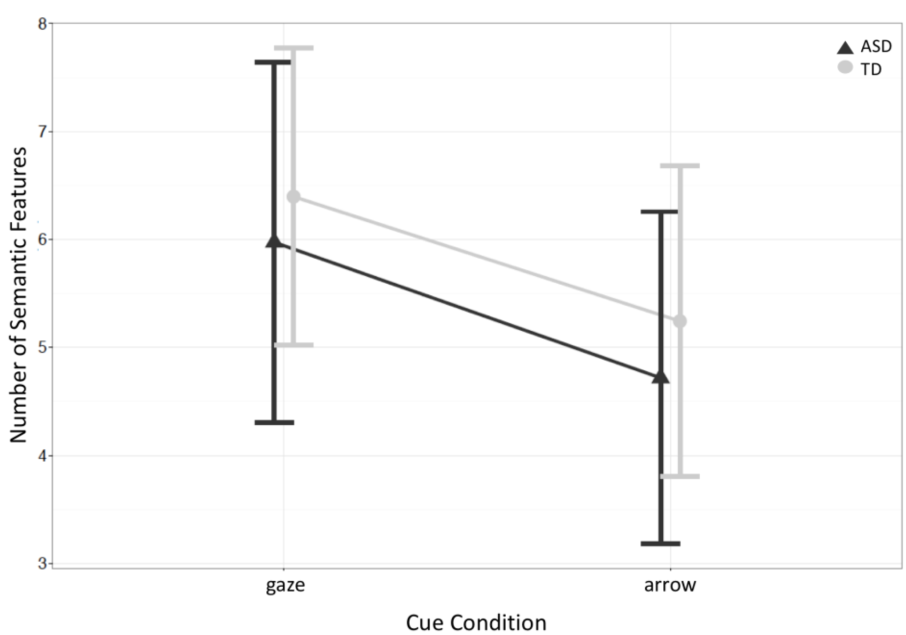
Target words - number of semantic features. Next, I examined the number of semantic features children recalled in their valid target object descriptions. Four additional children were excluded when they never provided valid target object descriptions at either visit, despite having provided valid familiar object descriptions (3 ASD, 1 TD). Therefore, the number of children included in this analyses included 18 children with ASD and 23 TD children.

A linear mixed model was specified with a random intercept of participant and fixed effects of group (ASD, TD), cue condition (gaze, arrow), and visit (visit 1, visit 2). Results differed with and without outliers. With all children, there was a marginal effect of cue condition, $F(1, 128.65) = 3.24$, $p = .07$, and marginal group by visit interaction, $F(1, 128.68) = 2.92$, $p = .09$. All other main effects (group, visit) and interactions (i.e., group by cue condition, cue condition by visit, and the three-way interaction of group, cue condition, and visit) were not significant (Fs < 2.38, ps > .13). The marginal cue condition effect demonstrated that children were recalling more features with gaze (*PMM* = 5.88, SE = .50) than with an arrow (*PMM* = 5.01, SE = .51). The interaction indicated that while children with ASD did not differ in their recollection of features between visits (visit 1 *PMM* = 5.02, *SE* = .66; visit 2 *PMM* = 5.10, *SE* .80), TD children were providing more features at visit 1 than visit 2 (visit 1 *PMM* = 6.62, *SE* = .61; visit 2 *PMM* = 5.05, *SE* = .79).

However, after removing outliers determined using Cook's distance (2 ASD), there was a significant main effect of cue condition, $F(1, 116.97) = 5.88$, $p = .02$, and a significant group by visit interaction, $F(1, 117.94) = 5.88$, $p = .02$. There were no other main effects or interactions (Fs < 1.84, ps > .18). The cue condition effect revealed that all children provided more semantic features in the gaze (*PMM* = 6.19, *SE* = .54) over the arrow condition (*PMM* = 4.98, *SE* = .53), *PMM* difference = 1.20, $d = .18$. A post hoc test of the group by visit interaction found no differences between groups of children at visit 1 or visit 2 (ps >.28). The significant interaction was due to TD children providing fewer semantic features at visit 2 than visit 1 (visit 1 *PMM* = 6.62, *SE* = .62; visit 2 *PMM* = 5.02, *SE* = .79), $t(130.67) = 2.36$, $p = .08$, *PMM* difference = 1.59, *d* = .21, although this finding was marginal and did not reach significance. This finding should be interpreted with caution because visit 2 included descriptions from a smaller subset of children (10 ASD, 15 TD), as noted above in the finding of fewer valid target object descriptions by visit 2. Children with ASD did not differ in the number of semantic features recalled between visit 1 and 2 (visit 1 *PMM* = 4.93, *SE* = .71; visit 2 *PMM* = 5.77, *SE* = .89), $t(130.76) = -1.13$, $p = .70$, *PMM* difference = -.84, *d* = -.20.

These findings exhibit that although TD children tended to remember fewer semantic features at their second visit, across visits, both children with ASD and TD children provided more semantic features when learning with gaze versus an arrow cue. Figure 13 depicts the cue condition effect in each group of children in each visit collapsed across visits. Table 12 provides the raw number of semantic features by cue condition, group, and visit.

Figure 13. Word Description: Predicted Marginal Means for the Number of Semantic Features in Each Group of Children Collapsed Across Visits



Points represent predicted marginal means and error bars are 95% confidence intervals.

Table 12. Word Description: Number of Semantic Features Provided for Target Objects – Removing Outliers

Visit 1	ASD (*n* = 16)	TD (*n* = 23)	*d*
Gaze	5.00 (3.20)	7.04 (2.20)	-.75
	0 - 14	2 - 15	
Arrow	5.17 (2.45)	6.39 (2.70)	-.47
	2 - 14	2 - 16	
d	-.06	.26	
Visit 2	ASD (*n* = 10)	TD (*n* = 15)	
Gaze	7.29 (3.33)	6.50 (2.74)	.27
	1 - 13	1 - 13	
Arrow	4.92 (1.94)	5.33 (1.45)	-.23
	2 - 10	2 - 12	
d	.95	.50	

The values shown are the mean(*SD*) and range. Negative Cohen's *d* values represent higher scores for typically developing children when comparing across groups and higher scores for the arrow when comparing across cue conditions.

Word generalization. After word descriptions, children were asked if the target label extended to different exemplars of the target object. In contrast to word association and word description which directly measured children's semantic knowledge about the target object, word generalization measured knowledge about a category of objects referred to by the label. I examined the accuracy of extensions and response times. In practice trials, all children demonstrated that they could extend a known label (hammer) to different exemplars of a hammer (photograph of a hammer, line drawing of a hammer) and reject a non-exemplar (photograph of scissors). Therefore, all children demonstrated that they understood the procedure. However, the procedure was revised partway through the study, thus data is only available for 21 children with ASD and 17 TD children who participated after the revised procedure.

Target words – accuracy. Please see Figure 5 for images of exemplars. For each target object (2 per cue condition), a trial included an image of the original target and its exemplars (6 maximum images), an image of the original distractor and its exemplars (3 maximum images), and images of a familiar object (2 maximum images, e.g., a hammer). Any trials where children were unable to correctly reject the familiar object image were excluded, but this did not result in removing all data from any children.

Accuracy scores included one point for correctly extending the label to each exemplar of the target object (6 maximum) and correctly rejecting each exemplar of the distractor object (3 maximum), for a total possible range of 0 to 9 correct scores per target object. The linear mixed model was specified with a random intercept of participant and fixed factors of group (ASD, TD), cue condition (gaze, arrow) and visit (visit 1, visit 2). There were no significant main effects or interactions between group, cue condition or visit (Fs < 2.80, ps ≥ .10). Table 13 demonstrates that in both groups accuracy was high in gaze and arrow conditions, with mean scores around 6 to 7, though distributions were skewed since the median was often 8; there was wide variation that covered the full range of 0 to 9 for accuracy scores. When children failed to extend the target label to an exemplar, this occurred most with the different shape exemplar (28% of failed extensions). Failed extensions for other exemplars were observed at similar rates across exemplars: original (14%), black and white (14%), color (14%), drawings of full objects (13%), and drawings of the object in parts (17%). Table 13 exhibits the accuracy scores by group, cue condition, and visit. Accuracy data from the word generalization task demonstrates that children in both groups were able to correctly generalize the target label to its exemplars and reject generalizing the label to distractor objects.

Table 13. Word Generalization: Accuracy Scores

Visit 1	ASD (*n* = 21)	TD (*n* = 17)	*d*
Gaze	6.24 (2.40)	6.36 (3.01)	-.05
	Mdn = 8	*Mdn* = 8	
Arrow	6.30 (2.67)	7.10 (3.33)	-.27
	Mdn = 8	*Mdn* = 9	
d	-.02	-.23	
Visit 2			
Gaze	6.28 (2.45)	5.97 (3.08)	.11
	Mdn = 8	*Mdn* = 8	
Arrow	6.32 (2.73)	6.97 (3.08)	-.23
	Mdn = 8	*Mdn* = 8.5	
d	-.01	-.32	

The values shown are the mean(*SD*) and median. Ranges were 0 to 9 for all subgroups. Negative Cohen's *d* values represent higher scores for typically developing children when comparing across groups and higher scores for arrow when comparing across cue conditions.

Target words – response times. In addition to evaluating the accuracy of children's extensions, response times were analyzed when children correctly extended the label to the target object. Response times were not analyzed for children's correct rejections to the distractor object or familiar objects to limit the number of fixed effects in the model. Accurate selections of the target object were only included in trials where 1) children correctly rejected familiar objects (as seen above in accuracy scores), as well as when 2) children correctly rejected distractor objects. These criteria better ensured that children's response times in the trial reflected a more precise understanding of the target object, which included rejecting images of the familiar objects as well as the distractor objects. This data cleaning resulted in data from 19 children with ASD and 17 TD children (from 21 ASD, 17 TD). A maximal linear mixed model was specified, because including random slopes resulted in a significantly better fit over the parsimonious model of only a random intercept of participant. A random intercept of exemplar type (levels: original, black and white, different color, different shape, drawing of full object, drawing of object in parts) was also included since there were some differences in accuracy depending on the exemplar (as seen above in accuracy results). Exemplar type was not included as a fixed effect since it was not part of the research questions.

A linear mixed model was specified with a random intercept of participant and random slopes of cue condition and visit (including interactions between cue condition and visit), a random intercept of group, and a random intercept of exemplar type. Fixed effects included those of group (ASD, TD), cue condition (gaze, arrow), and visit (visit 1, visit 2). Response times were log transformed due to heteroscedasticity. There was only a significant main effect of visit, $F(1, 34.41) = 8.60$, $p = .01$, and no other significant main effects or interactions between group, cue condition, or visit (Fs < 2.10, ps $> .16$). All children responded significantly faster at the second visit (*PMM* = 1367.10, *SE* = 572.63) than the first visit (*PMM* = 1545.36, *SE* = 647.39), *PMM* difference = 178.26, $d = .07$. Table 14 provides the raw response times for children by cue condition, group, and visit.

Faster responses to target objects specifically at the second visit suggests that children remembered the target objects from the first visit, though this finding did not examine if children would have been faster with all objects and thus simply remembered how to do the task. There were no cue condition differences on either measure of accuracy or response times.

Table 14. Word Generalization: Response Times (ms)

Visit 1	ASD (n = 17)	TD (n = 17)	d
Gaze	1854.12 (1427.78)	1425.29 (789.11)	.36
	480 - 7487	486 - 4662	
Arrow	1915.36 (1569.36)	1600.85 (1144.39)	.23
	446 - 9615	447 - 9134	
d	-.04	-.18	
Visit 2	ASD (n = 19)	TD (n = 16)	
Gaze	1513.17 (1065.93)	1479.22 (1653.87)	.03
	473 - 9181	477 - 13221	
Arrow	1762.61 (1519.04)	1601.14 (1540.79)	.11
	372-11169	320 - 10981	
d	-.19	-.08	

The values shown are the mean(*SD*) and range. Negative Cohen's *d* values represent higher scores (slower responses) for typically developing children when comparing across groups and higher scores (slower responses) for arrow when comparing across cue conditions.

Word production. At the end of each testing session, children were asked to provide the name for different objects, for a measure of whether children could recall the name of the target object. Because there were so few correct responses, comparisons between cue conditions would not be informative. Instead, I used Fisher's exact test to compare the number of children in each group of children who did or did not provide a minimum of one correct label. All 24 children with ASD and 24 TD children from the matched group were included in word production analyses.

There were no significant differences in distributions between groups at visit 1, Fisher's exact test $p = .12$, or at visit 2, Fisher's exact test $p = .38$. At visit 1, the number of children who provided a minimum of one correct label was higher in children with ASD than TD children (11 ASD, 5 TD), though this increased in TD children by visit 2 (12 ASD, 8 TD). As seen in Table

15, the median number of correct productions at the first visit was 1 label for children with ASD and TD children. At the second visit, both groups of children slightly improved with a median of 2 labels for children with ASD and a median of 1.5 labels for TD children. Across all children and visits, 16% of responses were correct versus 84% of responses that were incorrect. Of the correct responses, 77% were fully correct labels (e.g., *pagoune*) and 23% were partially correct (e.g., *paloule* or *spagoune* for pagoune). Thus, on the word production task, less than a quarter of all children's responses were the correct label, with the majority of responses as the fully correct label. Over half of the correct label productions at visit 1 were for objects that were taught last (66%), indicating a recency effect. A similar recency effect was seen at visit 2 where almost half of correct label productions were for objects that were taught last during visit 1 (44%).

When children provided incorrect responses, 53% of incorrect responses were when children said that they didn't know the name of the object, 10% of responses were when children said the object did not have a name when it was a target object, 10% of responses were when children used real names to try and describe the target object (e.g., *balance beam* for pagoune), 8% of responses were when children made up nonsense labels (e.g., *tochi* for pagoune), 11% of the time children provided the wrong label for the target (e.g., the child said *mimole* when a pagoune was shown), 1% of the time children's responses were unclear because the phonemes were a combination of target labels (e.g., *mopen* could be fopam and mimole), and 7% of the time names were repeated for two or more objects.

Table 15. Word Production: The Number of Children's Correct and Incorrect Responses

Visit 1	ASD (*n* = 24)	TD (*n* = 24)
Correct	1.45 (.69); *Mdn* = 1	1.20 (.45); *Mdn* = 1
	1 – 3	1 – 2
Incorrect	3.33 (.87); *Mdn* = 4	3.75 (.53); *Mdn* = 4
	1 – 4	2 – 4
Visit 2		
Correct	2.00 (1.04); *Mdn* = 2	1.75 (.89); *Mdn* = 1.5
	1 – 4	1 – 3
Incorrect	3.13 (1.10); *Mdn* = 4	3.42 (.97); *Mdn* = 4
	1 – 4	1 – 4

The values shown are the mean(*SD*), median, and range.

Because children were also shown images of the distractor objects, I checked to see if children mis-mapped the target label to the distractor instead of the target. At visit 1, only 2 children with ASD and 6 TD children mis-mapped the target label to its distractor object at least once. At visit 2, these numbers increased with 7 children with ASD and 8 TD children who mis-mapped the target label to its distractor object at least once. Across all children and visits, 8% of responses to distractor objects were instances of a mis-mapped target label (93% fully correct target labels; 7% of the mis-mapped target label were partially correct). These instances of mis-mapping the target label to the distractor object indicates that these children did not use the cue to associate the label to the target object, but followed their own interests to associate the label to the likely more interesting distractor object.

When children did not mis-map the label to the distractor objects (92% of responses to distractor objects), the majority of responses were that the distractor objects did not have a name (technically the appropriate response for a distractor object; 37%), or that children did not know a name (also an appropriate response for a distractor object; 45%). In some cases, children again used a real name to try to describe the distractor object (8%), made up a nonsense name (4%), or provided an incorrect target label that was not a mis-mapping (6%, e.g., children said *mimole* when the distractor was not the distractor for the mimole). Thus, when children were shown distractor objects, the majority of the time children were not mis-mapping the target label to the distractor.

In summary, across all word learning measures, there were three differences between cue conditions that favored learning from a gaze over an arrow cue, and no differences were found on these measures between children with ASD and TD children. Children in both groups spent more time looking on the area of the gaze cue relative to the area of the control cue, children located the target faster during test in the gaze over the arrow condition, and children recalled more semantic features in the gaze over the arrow condition. A table of all results can be seen in Table 16.

Table 16. Word Learning: Summary of Results

Type of Measure	Measure	Difference between gaze versus arrow?
Visual Attention	Attention to the target object	✗
	Attention to the cue: Proportion looking time to cue area	✓
	Attention to the cue: Contingent looking to target object	✗
	Test: Latency to target object	✓
Learning	Test: Word Recognition (pointing)	✗
In-depth learning	Word Association *Note: floor effects*	✗ (grey)
	Word Description: Object identification	✗
	Word Description: Number of semantic features	✓
	Word Generalization (accuracy and response times) *Note: ceiling effects*	✗ (grey)
	Word Production *Note: floor effects*	✗ (grey)

This figure depicts a summary of the findings regarding cue condition differences. ✗ = no significant differences between gaze and arrow cues (shown in black), ✓ = significant differences between gaze and arrow cues (shown in green). ✗ = floor effects (word association, word production), or ceiling effects (word generalization), suggesting caution when interpreting these non-differences (shown in a light grey).

Consistency across in-depth learning measures. Finally, I wanted to qualitatively examine individual children's performances across in-depth learning measures. That is, how consistently do children provide an answer related to the target object across the in-depth learning measures? If children can demonstrate learning consistently across multiple measures, then this would demonstrate a more robust understanding of the label-object association that can be tapped into by using different measures. For example, describing the object and producing the name gives some sense that children understand that a particular object they can describe to some degree has a name, whereas simply describing the object but not being able to produce the name demonstrates a weaker understanding of the label-object association. Word generalization was not included because only a subset of the matched group underwent the revised protocol. It is important to note that excluding word generalization leaves us with only measures of expressive language, which appeared to be more difficult for children. All 24 children with ASD and 24 TD children were included.

Children's performance on each measure was converted into a binary score, because the measures were coded differently and a binary score provided a simple and similar way to assess learning across measures. A score of 1 was given if children identified the target (word description, word production) or in the case of word association, when children provided a semantic response. A score of 0 was given for all other responses (e.g., word association – clang response, word description – describing another object, word production – making up a name). Each child had four possible trials, or 4 possible targets to learn, with three different expressive language measures of in-depth learning per target object. If children demonstrated in-depth learning of the target across two or all three of the three measures, they were considered as trials with *consistent* learning of the target. In-depth learning with only one identification of the target was considered an *inconsistent* trial, and no identifications were considered as trials with *no learning*.

I used this measure to get a sense of how individual children were learning across measures, but because there were few instances of consistent learning per child, I divided children into three groups that separated those with any consistent learning (minimum of one target object), only inconsistent learning for all four target objects, or no learning for all four target objects. As seen in Table 17, while at visit 1 over half of children in both groups demonstrated consistent expressive learning for at least one target object, the number of children with a minimum of one trial with consistent learning decreased by visit 2. These descriptive findings exhibit that a similar number of both children with ASD and TD children demonstrated consistent learning (though this was on a minimum of one trial), inconsistent learning, and no learning in both visits, indicating that children with ASD demonstrated a similar pattern of responding to expressive language measures as TD children.

Table 17. Word Learning: Number of Children with Consistent, Inconsistent or No Learning Across Expressive Language Measures

Visit 1	ASD (n = 24)	TD (n = 24)
Consistent learning (minimum of one target object)	14	18
Inconsistent learning only	8	5
No learning only	2	1
Visit 2		
Consistent learning (minimum of one target object)	9	11
Inconsistent learning only	10	9
No learning only	5	4

DISCUSSION

In this chapter I examined how children with ASD and typically-developing children learn new words with referential gaze. Specifically, I compared a directional arrow cue to a potentially intentional referential gaze cue to determine if gaze was interpreted as intentional. I also tested the potential implications of intention understanding on both word recognition and in-depth word learning measures. Prior work on referential gaze has demonstrated that it may do more than simply enhance attentional salience (Aldaqre et al., 2015; Field, 2016). The one study (Field, 2016) that has directly compared an arrow and a gaze cue had cues that differed on multiple features (size, color, movement) and tested different children with each cue, which made it difficult to ascribe any differences in learning between cues to an intentional reading of referential gaze.

Arrows are a good match for gaze in that both are familiar and both are used as signs of direction, yet they potentially differ on intentionality. I matched arrow and gaze cue conditions on features of the cues such as the size, color, duration and angle of motion, and relative distance from the target objects. Additionally, the same actress was present in both conditions and children heard the same labels with both cues. With this stringent comparison, any differences between the cues could be explained by an important theoretical difference between an arrow and a referential gaze cue: the proposed intentionality of gaze. The measures used to assess potential differences between cues included children’s visual attention to different parts of a scene (e.g., the target object and the area of the cue), as well as in-depth learning measures, conducted both immediately after teaching and one week later. In prior studies, the performance of typically-developing children has been used as a baseline to support the presence of intention understanding (Baldwin, 1993b; Bani Hani et al., 2012) or the lack of intention understanding in children with ASD (Norbury et al., 2010). However, these studies did not include a control for referential gaze, which limits our understanding of whether either children with ASD or typically-developing children learn because they treat referential gaze as an intentional cue or if there are other explanations, such as treating referential gaze as a directional cue. The current study is novel in comparing performance on measures of visual attention and in-depth learning between a directional arrow cue and a potentially intentional referential gaze cue.

The results of this dissertation identified three key differences between an arrow and a referential gaze cue on visual attention and in-depth learning measures. While each of these differences alone would be weak evidence to support an intentional reading of gaze, the combination of all three findings suggest that children may not simply treat referential gaze as a directional cue. These three differences provide evidence that suggests that referential gaze may in fact be treated as an indicator of underlying intention in both children with ASD and typically-developing children. First, children in both groups spent a greater proportion of looking time at the area of a gaze cue versus the area of an arrow cue. Second, children in both groups located a target faster during test in the gaze condition versus the arrow condition. Finally, children in both groups recalled more semantic features in the gaze condition versus the arrow condition. Despite some caveats to these findings (discussed below), they provide evidence that support a rich interpretation of referential gaze in typically-developing children in a word learning context (e.g., Baldwin, 1993b; Moore et al. 1999; Norbury et al., 2010). Notably, these findings demonstrate also that individuals with ASD are capable of an intentional reading of gaze for successful word learning (Aldaqre et al., 2015; Bani Hani et al., 2012). Moreover, the finding of increased recollection of semantic features after learning with a referential gaze cue relative to an arrow cue suggests that intention reading may be positively related to children's in-depth learning. This discussion is separated into two sections, where the first section discusses the three key differences found between the arrow and referential gaze conditions, and the second section discusses what can be determined about word learning in general between children with ASD and typically-developing children.

Perhaps surprisingly, no differences between groups of children were noted on any of the measures included in this study. This lack of significant differences between groups could suggest that as seen in prior word learning studies, when groups are well matched on factors such as nonverbal IQ and chronological age, children with ASD are able to perform similarly to typically-developing children. However, null findings could also be due to a difference between groups that could not be detected in this sample. While a longer discussion of between-group differences can be seen in the second section of this discussion, the first section notes when other studies also found no significant differences between groups on each respective measure.

Children with ASD and TD Children Treat Gaze as Intentional: Three Findings from Word Learning

In the gaze condition, children spent a greater proportion of looking time at the area of the cue. The first key finding was that both groups of children spent a greater proportion of looking time at the area of the gaze cue versus the area of the control cue. This finding was a main effect across baseline and teaching phases, thus attention in the gaze condition is collapsed across attention to direct gaze and referential gaze (area of the gaze cue), and attention in the arrow condition is collapsed across attention to the static circle and the arrow (area of the control cue). Because children were looking less at the area of the control cue relative to the gaze cue, this could have affected how they followed the arrow cue during teaching. However, two findings suggest that less looking at the area of the control cue did not affect how they followed the arrow cue. First, a video phase main effect indicated that for both cues, children looked at the area of the cue significantly more during teaching than baseline, demonstrating that children noticed that both cues were different during teaching. Second, during teaching, children did not differ in their attention to the target object between cue conditions, demonstrating that children followed both cues to the target object. Thus although children in both groups looked longer at a gaze cue relative to a control cue, this did not affect their ability to follow both cues to target objects during teaching. There were no differences between groups in their visual attention to the cue or to the target object, which is similar to other word learning studies (Akechi et al., 2011; 2013; Norbury et al., 2010; Tenenbaum et al., 2014).

However, these results demonstrate that during baseline, children were already spending a greater proportion of looking time at the area of the gaze cue relative to the area of the control cue. Thus, more attention during teaching could have resulted from an initial interest in looking at the area of the gaze cue relative to the control cue during baseline. One the one hand, a greater proportion of looking time at the area of the gaze cue could still indicate that both groups of children read gaze as an intentional cue (Csibra & Gergeley, 2009). On the other hand, children could be looking more at the area of the gaze cue during baseline because direct gaze is more perceptually salient than the area of the control cue during baseline which was a static circle. In other words, if children were drawn to the salience of direct gaze during baseline, it would not be surprising that this relative difference might have continued during teaching. These baseline differences make it difficult to resolve the central question of whether a directional arrow is used similarly or differently from a referential gaze cue.

The finding of a greater proportion of looking time to the area of gaze versus the area of the control cue alone would not be able to indicate whether gaze is an intentional cue. Future studies need to address whether differences during baseline attention to the cue can affect children's attention to cues during teaching. Other potential reasons for differences between cues include the placement of the cues (e.g., gaze embedded in a face versus a control cue above her head), or perceptual complexity (i.e., gaze is a more complex cue with variations in color and texture versus the black and white control cue, luminance of the iris and sclera; Doherty et al., 2015). However, considering this finding in combination with other findings of cue condition differences discussed below (e.g., latency, word description), the evidence appears to be consistent with the interpretation that children with ASD or typical development may be treating referential gaze as intentional.

Despite cue condition differences in the proportion of looking time to the area of the cue, there were no significant cue condition differences on the measure of contingent looking difference. Contingent looking between a person's face and a target object has been used to indicate that children are sensitive to the relationship between gaze and the target object, with some authors proposing that this sensitivity is indicative of understanding another's intentions (Baldwin, 1993b; Norbury et al., 2010). Prior studies have suggested that children with ASD do not treat referential gaze as intentional because they provided less contingent looking relative to typically-developing children (Gillespie-Lynch et al., 2013; Norbury et al., 2010; Preissler & Carey, 2005). However, in all of these studies, there has not been a control cue for referential gaze. Therefore, these studies have assumed that contingent looking between gaze and a target object is intentional, without examining whether contingent looking can occur with any cue, and whether any relative difficulties in children with ASD could exist with any cue or only a referential gaze cue. If children engage in contingent looking differently when the cue is gaze relative to other non-intentional cues, then this measure would better indicate an intentional reading of gaze. Furthermore, if difficulties in contingent looking were found in children with ASD with only referential gaze and not with other cues, then it would be more convincing that less contingent looking would be because of an inability to treat gaze as intentional in children with ASD.

In this study, both groups of children provided more contingent looking with a referential gaze cue relative to an arrow cue, though this was not significant ($p = .08$). It is possible that the lack of a significant effect could be a result of power, because many children provided few contingent looks and 6 children were removed for not providing any contingent looks (3 ASD, 3 TD). Thus it remains to be seen if contingent looking may be a measure that can reveal intention understanding in both groups (Baldwin, 1993b). While other studies have found significant differences in contingent looking between children with ASD and typically-developing children (Gillespie-Lynch et al., 2013; Norbury et al., 2010; Preissler & Carey, 2005), the lack of significant differences between groups in the current study could be due to how the measures were collected. For example, the current study used fixations on the specific area of gaze versus the area of the speaker's face. It could be that contingent looking does not differ between groups when contingent looking is specified to be the area around the eyes, whereas there may be group differences when considering the area of the face more generally. Nevertheless, based on the non-significant finding between cue conditions in this sample, the findings from this measure are not considered as evidence that can be used to support an intentional reading of gaze.

In the gaze condition, children located the target faster during test. The second difference between cue conditions was that both children with ASD and typically-developing children were faster to locate the target during test in the gaze versus the arrow condition. No differences were found between groups, which is in line with Tenenbaum and colleagues (2014), who, thus far, was the only other study to examine this measure in a word learning context in children with ASD. During test, neither of the cues were present, thus there was nothing else to look at except the target and the distractor objects. On average, children fixated 150 ms faster on the target object in the gaze condition, where it took approximately 750 ms to fixate on the target in the gaze condition relative to approximately 900 ms in the arrow condition (standard errors in both conditions were around 50 ms). This difference was seen after 4 children with ASD were excluded because they were identified as potential outliers based on Cook's distance.

One possible reason for this cue condition difference in latency could simply be because of differential attention to the cues prior to the test phase. I found that children spent a greater proportion of time looking at the area of gaze relative to the area of the control cue, which may have influenced latencies during test. Yet latencies were recorded only with the objects present in the scene and not the cues, thus latencies would be less affected by differential attention to the cue prior to the test phase. Instead, latencies to the target object may have been affected if there was differential attention to the target objects between cue conditions. Measures that can speak to children's attention to the target object include their target advantage scores and contingent looking difference, neither of which resulted in significant differences between cue conditions. Therefore, while cue condition differences in latency could be due to differences in attention to the area of the cue, the lack of differences in attention to the target object, which is the area of interest used in latency, suggests that another reason for this cue condition difference could be explained by an intentional reading of gaze.

In the gaze condition, children recalled more semantic features. Norbury et al. (2010) proposed that because children with ASD relative to typically-developing children spent less time attending to gaze in a word learning context, that children with ASD may have a less developed understanding of gaze as an intentional cue. Moreover, they found that children with ASD recalled a smaller proportion of semantic features than typically developing children. In this study I sought to test if there was a link between an intentional reading of gaze and children's recollection of semantic features. To examine this link, I compared children's descriptions of novel words when learning with a directional arrow cue or a potentially intentional referential gaze cue and counted the number of semantic features that children recalled about the target object. Semantic features referred to physical features of objects (e.g., *yellow, cylinder*), as well as functions (e.g., *push the button*) and uniquely identified the target object relative to other objects. In contrast to the other word learning measures in this study (e.g., word recognition, word generalization, word production), children's descriptions provided a measure of children's knowledge of the referent that did not rely on seeing an image of the target. The third and final cue condition difference was found in children's in-depth learning, where both children with ASD and typically-developing children recalled more semantic features in the gaze versus the arrow condition across both visits.

Importantly, the number of semantic features provided by children were only examined for descriptions where a coder could first identify that the child was describing the target object. The target object identifications were compared between groups of children using the levels of success at learning, and results revealed two main findings to keep in mind when considering the finding in semantic features. First, the difference in the number of semantic features was not simply due to more target objects identified in the gaze relative to the arrow condition. At both visit 1 and visit 2, 14% or less (3 children or fewer) in each group of children had success with only the gaze or the arrow cue at both visits. Instead, more children in both groups had success with both cues (those that were able to describe the target on three or all four trials), with around 50% of children at visit 1 in both groups. Another point to keep in mind is that one week later, there were fewer target object identifications in both groups. The number of children who had success with both cues dropped from around 50% in both groups to 14% (3/21) of children with ASD and 4% (1/24) of typically-developing children. Thus, though the main effect of cue condition takes into account both visits, more of these descriptions were from the first rather than the second visit. At the second visit, it was either more difficult for children to remember target objects and/or descriptions were less clear or less detailed such that the coder was unable to identify descriptions as target objects.

For the number of semantic features, the difference between gaze and arrow conditions appears to be small, with around 1 semantic feature recalled more in the gaze versus the arrow condition, where in the gaze condition children recalled around 6 features in contrast to around 5 features in the arrow condition (standard errors in both conditions were less than 1). Though this difference is small, the fact that there was a significant difference is surprising given that children were not explicitly told to pay attention to these features, experimenters were blind when interacting with children, and the coder was blind to the cue condition. This difference could not have been due to differential attention to the target object during teaching, because there were no significant differences in target advantage scores during baseline, teaching, or test phases. Therefore, this third and final cue condition difference contributes to a pattern of differences between gaze and arrow conditions, suggesting that both children with ASD and typically-developing children may be treating gaze as an intentional cue. Moreover, in both groups, there may be important positive implications of intention understanding where there may be a real advantage for recalling semantic features when learning with a referential gaze cue.

Another key point to keep in mind about this finding is that in the matched sample, 8 children with ASD and 1 typically-developing child were not included because they could not describe familiar objects (3 ASD), they never described the target object at both visits (3 ASD, 1 TD), or were deemed outliers using Cook's distance (2 ASD). This resulted in 16 of 24 children with ASD and 23 of 24 typically-developing children providing data to this finding. Thus, aside from the two children who were removed as influential points using Cook's distance, these findings take into account only those children who were able to provide valid familiar object and target object descriptions at both visits.

Aside from the possibility of whether they indicate intention or not, there are other differences between gaze and arrow cues as noted above. One other important difference between these cues is familiarity. Rather than a completely novel cue, I chose an arrow because it was likely to be a familiar directional cue, which is similar to referential gaze. However, referential gaze is more than just a familiar cue, it is a cue that children are exposed to from birth, and respond to as early as 4 months old (e.g., Farroni, Johnson, Brockbank, & Simion, 2000). Thus, we cannot distinguish the separate roles of an intentional reading of referential gaze or the familiarity with referential gaze which children build with every interaction they have from birth. Nevertheless, because arrow and gaze cues were well matched on size, motion and physical features, and the primary difference between these cue conditions is likely to be the proposed intentionality of gaze (Baldwin, 1993b; Baron-Cohen et al., 1997; Csibra & Geregely, 2009; Meltzoff, 2007), I have chosen to take a rich interpretation of the data which appear to be consistent with the explanation that gaze is different from a directional arrow cue because gaze is an intentional cue.

There were no significant differences between groups of children despite weaker expressive language abilities in children with ASD, which is contrary to Norbury and colleagues (2010) whose sample also included children with ASD with weaker expressive language abilities, but found that children with ASD provided fewer semantic features than typically-developing children. Instead, children with ASD in this sample did not differ from typically-developing children in recalling semantic features. The lack of differences in recalling semantic information between children with ASD and typically-developing children is in line with other studies that examined descriptions with familiar (McGregor et al., 2012) or novel words (Gladfelter, 2014; Kreger, 2016) when children were matched on expressive vocabulary skills.

One reason for the lack of group differences in the current study could be because the word learning paradigm was different from that of Norbury and colleagues. While in the current study, children's learning was tested immediately after viewing each target object video, in Norbury et al. children were tested after responding to a set of videos including all target labels. Therefore, responding in the current study may have been easier for children with ASD because they could respond to one target label at a time. Another difference between studies is in what they were designed to answer. Norbury and colleagues asked children to watch and respond to multiple trials of situations when an actor provided referential gaze in contrast to when she provided direct gaze (which was uninformative about which object was the referent). Thus, children's descriptions in Norbury et al. were a result of a word learning scenario that included both uninformative and informative cues, which could have negatively impacted how well children with ASD recalled semantic features of the object indicated by the informative cue.

Therefore, in the current study, it is possible that groups did not differ because learning was assessed immediately after teaching each novel label and children participated in teaching scenarios where all trials were informative.

In summary, while there are caveats to each of the three cue condition differences noted above, the pattern of findings together suggests that these differences are consistent with gaze being treated as an intentional cue in contrast to a simply directional cue. Importantly, this intentional reading of gaze may be related to children recalling more semantic features about an object they learned. While prior studies with children with ASD have interpreted between-group differences as evidence for the lack of or presence of intention understanding in children with ASD (Baron-Cohen et al., 1997; Norbury et al., 2010; Preissler & Carey, 2005), I examined intention understanding in children with ASD by directly comparing children's performance between an arrow versus a referential gaze cue. No cue condition differences were seen on word recognition or other in-depth learning measures of word association, word generalization, and word production. The lack of differences may be because of floor and ceiling effects on these measures that made them less sensitive to differences between cues, if they did exist. Further considerations of between-group differences on these measures are discussed below. Overall, in the word learning context, an intentional reading of gaze may in fact be an important signal to children that they need to pay special attention to what comes after the cue because it may benefit their learning (Csibra & Gergely, 2009), and may have important benefits in how they process words and objects (Becchio et al., 2008; Waxman & Gelman, 2009). Most importantly, this intentional reading of gaze exists not only in typically-developing children, but also children with ASD.

The next section discusses in more detail the performance between groups of children on the word learning measures.

Word Learning in Children with ASD and Typically-developing Children

One somewhat surprising finding is that there were no group differences on any word learning measure, despite the fact that children with ASD had significantly weaker scores on standardized language measures relative to typically-developing children. Thus, if standardized language measures are an indication of children’s overall language abilities, then one might expect that children with ASD would have difficulty on the experimental word learning measures relative to typically-developing peers with stronger language abilities.

One reason for the lack of group differences could be that the tasks in this study required minimal receptive and expressive language skills in comparison to the standardized language measures used to assess children. For example, the CELF-4 language measures asked children to identify analogous words (Word Classes – Receptive), and verbally describe why those two words were analogous (Word Classes - Expressive). In contrast, word learning measures in the current study were much simpler. Receptive language measures in the current study included pointing (word recognition), or pressing one of two buttons (word generalization), while expressive language measures included providing a one-word association (word association), describing what they remember rather than needing to explain analogies (word description), and providing names to images (word production). Moreover, the amount of exposure to novel words used in this study was similar for children, whereas children in each group may have had different levels of exposure to familiar words used in the standardized language measures. Thus, children with ASD may have exhibited similar word recognition and in-depth learning as typically-developing children because the word learning tasks were simpler than the standardized language measures and children had the same exposure to the words used in the study.

The first explicit measure of word learning was word recognition, where children were asked to point to the target object at the end of the video. This measure was only conducted at visit 1, and there were no significant differences between groups in the distribution of their levels of success at learning. Results indicated that the majority of children with ASD and typically-developing children had success with both cues (71% ASD, 62% TD), though a few more children with ASD had limited or no success (25% ASD, 12% TD). These findings demonstrate that regardless of the cue, the majority of children in both groups were able to disengage from a distractor object (as seen by negative target advantage scores during baseline), and follow a cue (with minimal other information; e.g., no head turn, no body turn), to point to the target object during test. Thus these findings are in line with others that have demonstrated that children with ASD can learn new words when they need to break from their own interest and follow into another's focus of attention (Akechi et al., 2011; 2013; Bani Hani et al., 2012; Field, 2016; Gillespie-Lynch et al., 2013; Luyster & Lord, 2009; McGregor, Rost, Arenas, Farris-Trimble, & Stiles, 2013b).

However, other tasks, such as word association and word production resulted in floor effects, meaning that they appeared to be difficult for both groups of children. The word association task were included as a new exploratory measure to examine children's semantic knowledge (Sheng & McGregor, 2010). Yet this measure proved challenging for both children and the coder. With only a single word response, it was difficult for a coder to assess what the child was thinking, and some younger children had difficulty responding in this task. It is possible that more of children's responses were related to the target object (semantic response) than was coded, but for a coder to credit the child as providing a semantic response, the response needed to be one that could be associated to that specific target object (e.g., *yellow* only referred to the "pagoune," whereas the word *paper* could refer to multiple target objects). It is possible that more liberal coding criteria could increase the number of responses to more than the average of around 1 to 2 responses. An idea of what this would like could be seen if I combined across coding groups of semantic and any experimental object, which are the coding groups that are related to the objects in general. However, the average number of responses in the coding group of any experimental object was around .5, which would not increase responses much in either group.

Similarly, word production was difficult for both groups of children. At visit 1, the median value of correct responses in either group was around 1 correct label out of 4 possible. There did appear to be some improvement one week later in both groups, with a median of 2 labels in children with ASD and 1.5 labels in typically-developing children. Difficulty with word production is in line with other studies with children with ASD and typically-developing children (Walton & Ingersoll, 2013). Yet one notable difference between groups was that over double the number of children with ASD provided a minimum of one correct label at visit 1, in contrast to typically-developing children (11 ASD, 5 TD). This difference was reduced at visit 2, but still favored children with ASD (12 ASD, 8 TD), though at each visit the distribution of children who did or did not produce a minimum of one correct label was not significant between groups. These results are a similar pattern to Norbury et al. (2010), who asked children to name the novel object but examined the proportion of correct phonemes of the novel label. The authors found that children with ASD outperformed children with typical development in the proportion of phonemes produced at visit 1, with group differences diminishing four weeks later. This is in line with accounts of relatively advanced expressive over receptive language (i.e., speaking without understanding meaning), and preserved or enhanced short term auditory memory in children with ASD relative to typically-developing children (Nadig & Mulligan, 2016). Nevertheless, in the current study, there were no significant differences between groups in the number of children who did or did not produce a minimum of one correct label. It remains to be seen if children with ASD may have a relative strength in recalling novel labels, especially immediately after learning them.

In contrast to word association and word production, word generalization may have been too easy for children to detect a difference between cue conditions. Accuracy was on average high in both groups, both cue conditions, and at both visits, with positively skewed distributions of median scores around 8, and averages around 6 to 7, out of a possible score of 9. These high scores suggest that overall children were able to accurately extend the target label to other exemplars of the target, and accurately reject the distractor object and its exemplars. This is in line with other studies that have demonstrated that both children with ASD and typically-developing children are able to generalize a label (Bani Hani et al., 2012; Benjamin et al., 2015). When children were accurately extending the target label to the target object exemplars, children's reaction times did not differ between groups or cue conditions, though children were significantly faster at responding to target objects in visit 2 in comparison to visit 1. However, because analyses were only conducted on target objects, it is unclear whether faster response times are indicative of being faster to target objects specifically or if children would be faster at recognizing all target, distractor, and familiar objects one week later. To limit the number of fixed effects in the model, I did not include accurate responses to distractor and familiar objects (i.e., rejecting distractor and familiar objects). Moreover, the number of possible target (6), distractor (3), and familiar objects (2) differ, which would also make relative comparisons of response times between visits difficult to interpret. Therefore, while children are faster at responding to target objects at the second visit, whether this would occur generally to all objects or target objects specifically is unclear.

Finally, groups appeared to be similar at both visits in the consistency of children's performance across in-depth word learning measures requiring expressive language. One consistent learning trial was defined by children identifying the specific target on two or three measures out of three expressive word learning measures, with four possible words at each visit. Examining the consistency across measures provides some sense of the robustness of children's understanding of the label-object association if they can demonstrate learning of the target object across multiple measures. At visit 1, over half of children in both groups demonstrated at least one consistent learning trial (58% ASD, 75% TD), which decreased at visit 2 for both groups (38% ASD, 46% TD). Thus many children with ASD were able to demonstrate consistent learning at similar rates as typically-developing children, though the number of children in both groups decreased one week later. These results indicate that for a minimum of one target object, many children appeared to have a more robust understanding of the label-object association by demonstrating learning on two or three different measures. There was also no notable difference between groups in the number of children that demonstrated inconsistent learning (i.e., target was identified on one of the three measures), or no learning (i.e., target was not identified for any expressive word learning measure), which exhibits that it was not only children with ASD who had difficulty with consistent learning, but that there were both children with ASD and typically-developing children that demonstrated only inconsistent or no learning.

This survey of word learning measures demonstrates few differences between groups at both visits. Because of floor and ceiling effects, I am not able to determine if there are group differences. It could also be that there are no true group or cue condition differences on these tasks, and making tasks easier or harder, or adding more trials, would not make these findings any different. Yet what these floor and ceiling effects do indicate is that both groups had areas of difficulties and strengths on the different word learning measures. Future studies interested in using these measures will be able to better speak to the utility of these measures in assessing children's word learning. Word description, on the other hand, did not demonstrate clear floor or ceiling effects, suggesting that in the current study, this measure was the most sensitive to detect variation in children's in-depth learning. Moreover, for a minimum of one target word, at both visits there were children with ASD and typically-developing children who could demonstrate consistent learning across expressive language measures.

Conclusion

In conclusion, the findings from this study suggest that children in both groups do treat referential gaze as an intentional cue to contribute to their word learning. Moreover, in both groups, an intentional reading of gaze may have positive implications on the number of features recalled about an object. These findings support others who suggest that intention understanding is not impaired in all children with ASD (Aldaqre et al., 2015; Bani Hani et al., 2012), and can be seen in at least a subset of children of this heterogeneous population. Children with ASD also appeared to demonstrate similar visual attention to the video and word learning as typically-developing children when matched on chronological age, nonverbal IQ, ratio of girls to boys, and parental education. It is important for future studies to address other potential explanations for the effectiveness of referential gaze, such as the role of familiarity, and include a larger and/or a wider range of children with ASD to represent the full spectrum of the disorder (e.g., nonverbal IQ below the normal range).

CHAPTER 2: REFERENTIAL GAZE AND ACTION PREDICTION

INTRODUCTION

Referential gaze has garnered much attention in the context of word learning, but few studies have explored the role of referential gaze in other contexts, such as action prediction (Vivanti et al., 2011; Vivanti, Trembath, & Dissanayake, 2014). An example of how referential gaze was used in action prediction was seen in the General Introduction when Ben followed your referential gaze and put the brown block on the tower, rather than the flag that he was initially looking at. In contrast to the word learning context, in action prediction, children must predict how another person will act on an object, and demonstrate this act accordingly. Some authors have suggested that demonstrating this act means that children have reasoned what the person intends to do with that object (Meltzoff, 1995), which suggests that action prediction may be a potentially strong test of whether children use referential gaze as an intentional cue. Given the findings from Chapter 1, which suggest that children with ASD and typically-developing children do treat referential gaze as an intentional cue in the word learning context, the current study evaluated whether this intention understanding would also apply in the action prediction context.

The research suggests that children with ASD have difficulty using referential gaze to successfully predict another's actions. These studies have interpreted weaker performance in children with ASD, relative to children without ASD, as difficulty with understanding the intent behind referential gaze (Vivanti et al., 2011; 2014). Yet as seen in the word learning context, no studies to date have examined whether successful action prediction can also occur from following the directional cue of an arrow, rather than a potentially intentional cue of gaze. If both children can predict another's actions using a directional cue as well as they can when using a potentially intentional cue of referential gaze, this would suggest that learning in this context may not be due to intention understanding, but could be attributed to more lean processes such as attentional and spatial information (Heyes, 2014b; Leekam, 2016). If there are differences between cue conditions, then these could be due to an intentional reading of gaze.

Another possible way to examine intention understanding is by testing the proposal that children expect to learn something that can be generalized because they understand that gaze is an intentional cue (Csibra & Gergely, 2009). If children are able to generalize what they learn with a potentially intentional referential gaze cue and not an arrow cue, then this difference could be because children treated gaze as an intentional cue, where they expected to learn something that could be generalized (Csibra & Gergely, 2009). This generalization of children's learning has not yet been examined in the action prediction context.

Finally, a yet unstudied question is how individual children use cues across contexts. Word learning and action prediction contexts are different in what children are learning with cues (i.e., new words, predicting what someone will do), but the current study has presented the contexts similarly, using the same actress, the same cues, and the similar placement of information in the scene (e.g., target, distractor objects). Given that the presentation in these contexts are similar, it is important to examine if referential gaze or arrow cues are used similarly or differently across contexts. Comparing how children use these cues across contexts can help us better understand the role of a cue such as referential gaze, which is often used across multiple contexts in our everyday lives.

In the following sections I review the literature on action prediction in typical development as well as in children with ASD. This first section focuses on an experimental paradigm which launched much of the action prediction literature: Meltzoff's behavioral re-enactment procedure, also commonly referred to as the unfulfilled intentions paradigm (Meltzoff, 1995). Next, I review the studies that have examined the role of referential gaze in the action prediction context with children with or without ASD (Vivanti et al., 2011; 2014). This introduction will conclude with a description of the current study and research questions.

Action Prediction in Children with Typical Development or Autism Spectrum Disorder

Action prediction, or predicting an actor's goal (Aldridge, Stone, Sweeney, & Bower, 2000; Carpenter, Pennington, & Rogers, 2001; Vivanti et al., 2011), is one area of research within a larger scope of how children understand another's actions. This larger scope of action understanding covers areas such as the imitation of another's actions (for reviews on action understanding in general see Uithol & Paulus, 2013; Vivanti & Hamilton, 2014), predicting another's motor intentions (Becchio, Pierno, Mari, Lusher, & Castiello, 2007; Pierno, M, Glover, Georgiou, & Castiello, 2006), understanding irrational actions (Marsh & de C. Hamilton, 2011; Marsh, Pearson, Ropar, & de C. Hamilton, 2015), distinguishing intentional and accidental actions (Carpenter, Akhtar, & Tomasello, 1998; D'Entremont & Yazbek, 2006), as well as the focus of this study, action prediction. Much of the work on action prediction is based on whether children can complete an actor's unknown goal after watching an actor fail to complete their goal (e.g., pull sides off a dumbbell). For example, an actor would fail to pull off the sides of a dumbbell, then the child is given the dumbbell. If the child successfully takes off the sides, completing the goal is taken as understanding the actor's intent from the actor's failed demonstration (Meltzoff, 1995). This paradigm involves multiple body cues and is not specific to only a referential gaze cue, but I discuss them here because of their relevance to the context of action prediction. In contrast to word learning, where there were few studies that examined whether learning could occur from cues other than a potentially intentional cue of gaze, the action prediction literature in typical development has included control conditions to explore what information in the actor's failed attempts children may be using to predict another's unknown actions, and whether action prediction may result from intention understanding.

Meltzoff's (1995)'s behavioral re-enactment procedure revealed that 18-month-old children were able to predict the unknown end state of an actor's failed attempts on an object, the failed demonstration condition. Moreover, the performance of children in the failed demonstration condition did not differ from the performance of a different group of children, who saw the full demonstration including the end state, indicating that children could predict the same goal whether they saw a failed attempt or a full demonstration. In additional conditions, other possibilities were ruled out for how children could have predicted the unknown goal. In one condition, children were given the object without any demonstration (i.e., baseline condition), to examine if children could figure out the goal without any information. In another condition, the adult manipulated the object in a neutral manner without failed attempts (i.e., manipulation condition), to examine if watching an adult simply handling the object could help children predict the goal. In both the baseline and the manipulation conditions, children could not predict the end state as well as children in the failed or full demonstration conditions, demonstrating that predicting the actor's goal in failed and full demonstrations may have been due to children treating these situations differently, potentially because a person's failed attempts are treated as intentional.

Finally, in a second experiment, Meltzoff constructed a control condition that did not include a person acting on the dumbbell, but instead an inanimate object with mechanical arms acting on a dumbbell (though a human controlled the mechanical arms). The rationale was that if children had more difficulty with predicting the goal with the ‘failed attempts’ of an inanimate object than another group of children who saw a human’s failed attempts (using a similar motion as the inanimate object), then children who saw the human’s failed attempts are better because they are not simply attending to the mechanics of the actor’s movements, but because they understand the actor’s intent to complete the goal. Meltzoff found that children who saw the inanimate object’s failed attempts could not determine the end state as well as children who saw the human’s failed attempts. These findings, along with results from prior conditions (e.g., baseline, manipulation, full conditions), were interpreted as children being able to predict the end state because they understood a person’s intention from their failed attempts. The results of other researchers using a similar paradigm with multiple conditions (though only Meltzoff used an inanimate control) have concluded a rich interpretation that children can predict the end state by attributing an actor’s intent to his or her failed attempts (e.g., Bellagamba & Tomasello, 1999).

Yet in contrast to a rich interpretation, other researchers studying typical development (Huang, Heyes, and Charman 2002; 2006; Want & Harris, 2002) have challenged whether intention understanding is required for task success in the behavioral re-enactment procedure. One criticism is that the actor’s failed attempts bring attention to the physical affordances of the objects. For example, the failed attempt of trying to pull apart the sides of the dumbbell highlights that the sides might come off, or the failed attempt of trying to put a loop on a prong highlights that the loop could go on the prong. This highlighting is similar to the idea that referential gaze bringing attentional salience to a referent discussed in the word learning context (Aldaqre et al., 2015; Moore et al., 1999). Huang et al. (2002) showed that typically-developing children were able to complete the end state similarly in conditions with a full demonstration, a failed demonstration, and a condition where parts were presented to be spatially contiguous (e.g., holding the ends of the dumbbell or raising the loop to the end of the prong). Though Meltzoff also included a manipulation condition to control for the actor's attention to the object, Huang et al. focused on the spatial contiguity of the parts that led to the end state, rather than a general handling of the objects in Meltzoff. Therefore, the results in Huang et al. (2002) provide evidence that support a lean interpretation of how children treat an actor's failed attempts, where predicting the object-directed goals of an actor's actions may not require intention understanding.

This lean interpretation of how children treat an actor's actions suggests that if children with ASD are successful at predicting the actor's goal after watching a failed demonstration, that their success also may not be due to intention understanding. Yet studies have interpreted success in children with ASD as some degree of understanding of intent (Aldridge et al., 2000; Carpenter et al., 2001; Colombi et al., 2009), though the results of Carpenter and colleagues suggest otherwise. These authors included children with ASD or DD and found that both groups successfully completed the actor's goal from the actor's failed attempts, with no differences between groups. However, within children with ASD, their performance during the failed demonstration condition did not differ from other conditions of the full demonstration and manipulation conditions. Similar performance between the failed demonstration and the manipulation condition suggests that any type of object handling by the actor could have resulted in success for children with ASD. This argument is similar to the spatial contiguity condition conducted by Huang and colleagues (2002, 2006), and others have called this a "spotlight effect" of children simply attending to an actor's actions, meaning that success in the failed demonstration did not require understanding an actor's intent (D'Entremont & Yazbek, 2006). Yet Carpenter et al. still adopted a rich interpretation, suggesting perhaps a less complex level of intention understanding in children with ASD rather than a deficit (see Colombi et al., 2009 for a similar interpretation).

Given the difficulty with predicting children's use of objects designed to have a particular end goal, Parish-Morris and colleagues (2007) examined whether children could predict the goal of failed attempts of canonical (e.g., stacking rings on a plastic post) versus non-canonical actions (e.g., watering a truck) with children with ASD or typical development. Successfully predicting the actor's goal with non-canonical actions would be a stronger test than using objects that were designed to have an end goal (e.g., dumbbell with sides that come off), since determining the end state of a non-canonical action could less likely be attributed to an object's physical affordances (Huang et al., 2002; 2006). Two groups of children with typical development were matched to children with ASD, one group on chronological age and the other on language age. These authors found that children with ASD performed worse than both groups of typically-developing children with non-canonical actions. However, children with ASD performed worse than chronological-age matched peers on canonical sets, as well as non-canonical sets. Difficulty with both canonical and non-canonical sets in children with ASD suggests that there was no baseline level of task understanding, making it difficult to interpret their performance. Additionally, no other control conditions were conducted to assess whether the performance in typically-developing groups could be could be due to other factors such as general object handling, or spatial contiguity (Huang et al., 2002; 2006; Meltzoff, 1995).

Interestingly, Parish-Morris and colleagues (2007) included both word learning and action prediction contexts to examine intentionality in children with ASD and typically-developing children. They were interested in how potential intention understanding in children was related to their language and cognitive abilities, and found that children's vocabulary abilities were related to performance on tasks that the authors assumed to reflect intention understanding (e.g., discrepant word learning, non-canonical action prediction). However, these authors assumed that intention understanding was required in these contexts. Yet in both contexts it was unclear whether children were successful because they were treating the person's actions (e.g., providing gaze to an object to teach a name, attempting to complete a non-canonical action) as intentional or if children were learning using other lean information (e.g., the direction of gaze, the mechanics of the actor's attempts). Moreover, this examination across contexts does not provide information on whether the same children were as able to learn in the word learning context as they were in the action prediction context. This is important to note because if different children are demonstrating successful learning in one context and not the other, then learning in both contexts may not be due to the same process, such as intention understanding.

These studies in action prediction with the behavioral re-enactment procedure have provided both lean and rich interpretations of what it means when children can predict another's actions from their failed attempts. Some suggest that both children with ASD and typically-developing children treat failed attempts as the actor intending to convey their goal (Aldridge et al., 2000; Carpenter et al., 2001; Colombi et al., 2009; Meltzoff, 1995; Parish-Morris et al., 2007). Additionally, Meltzoff (1995) demonstrated that even when shown an inanimate object performing failed attempts and not a human's failed attempts, children can better predict the goal with a human, suggesting that at least typically-developing children may treat failed attempts as intentional. Yet others still take a lean interpretation, with evidence that typically-developing children can successfully predict another's goals by simply attending to the physical affordances of the object from the failed attempts (Huang et al., 2002; 2006; Want & Harris, 2002).

In a different variation of an action prediction context, two studies have examined how children with or without ASD treat the specific cue of referential gaze to predict another's unknown goal, with both studies taking a rich interpretation of how children treat referential gaze (Vivanti et al., 2011; 2014). They propose that intention understanding is intact in children who do not have ASD (i.e., children with typical development or mixed groups of children with DD and typical development), and weaker performance in children with ASD is because they lack intention understanding. However, these studies make the two assumptions noted earlier, where it is assumed that intention understanding is present in children who do not have ASD, and when children with ASD demonstrate a weaker performance relative to children who do not have ASD, that this is because children with ASD do not understand intent. Neither of these studies have tested whether successful action prediction (in their variation of this context manipulating referential gaze) can only occur with referential gaze, or if any other directional cue could also be used to predict an actor's goals in both children with or without ASD.

In the current study I examined intentionality in the context of action prediction in the same way it was tested in the context of word learning in Chapter 1: by comparing learning with a directional arrow cue versus a potentially intentional referential gaze cue. The same arrow and referential gaze cues were used as in Chapter 1, thus referential gaze and arrow cues were matched as closely as possible on size, color, and duration of movement, and any potential differences between cues is interpreted as an intentional reading of referential gaze. Moreover, I examined individual children's performance across contexts to investigate whether how children use cues is similar or different across contexts of word learning and action prediction. The following section will review the two studies that evaluated the role of referential gaze in the action prediction context with both children with or without ASD (Vivanti et al., 2011; 2014).

Action Prediction and Referential Gaze in ASD

Vivanti and colleagues (2011) adapted the Meltzoff unfulfilled intentions paradigm to investigate how children treated specific cues, such as an actor's head turn (including referential gaze), to predict the actor's unknown goal. They included 9- to 16-year-old children with ASD and typically-developing children (n = 18 in each group) matched on chronological age, language abilities, and performance IQ. In this adaptation, children watched a video of an actor stacking blue and red lego blocks in an alternating pattern (tower stacked with three blocks: red, blue, red), and the video froze before she reached for the next block, leaving a blue and a red lego block on the table. In the neutral condition, the actor remained still, thus children would likely predict the next block to be the one that continued the alternating sequence (i.e., the blue block). In the head-turning condition, the actor turned her head to the red block twice. This head turn thus served as a cue that the actor intended to break the alternating sequence and use the red block to build the tower. The experimenter set up the same physical blocks in front of children before the video, and after the video children completed the tower with the block they thought would go next. The authors measured children's visual attention using an eye tracker and whether children predicted the actor's actions on the physical blocks.

Findings revealed both similarities and differences between groups in their attention to the scene (i.e., actor's face, actor's action) and their action prediction abilities. Differences were seen in children's attention to the area of the actor's face and children's action prediction abilities. In both conditions, children with ASD provided fewer fixations to the actor's face than typically-developing children, and in the head-turning condition, children with ASD were less able than typically-developing children to choose the red block that would break the alternating sequence. Though the head-turning condition was difficult in both groups (averages for both groups were under 1, with a range of 0 to 2), the authors noted individual differences in that 10 of the 17 typically-developing children predicted that the actor would use the red block in at least 1 trial (i.e., the block indicated by the head-turn), whereas only 2 of 16 children with ASD chose the red block. There were no group differences in the neutral condition. Additionally, a positive correlation only in the ASD group suggests a relationship between more fixations to the actor's face in the head-turning condition and better action prediction abilities. Both groups provided a similar number of fixations to the area of the actor's actions (which included the area of the materials, though what parts of the body were included specifically was not mentioned) in both neutral and head-turning conditions, thus cue condition differences were less likely due to differences in attention to the actions and materials, but due to the manipulation of the head-turn. Based on the difficulty of children with ASD to predict the actor's goal in the head-turning condition, these authors concluded that children with ASD were unable to understand the actor's referential intent, as conveyed by the head turn.

However, though the study included a baseline condition (neutral condition) to see how children with ASD would perform without a cue, there was no condition that controlled for the head-turn cue. Therefore, these authors based their interpretation on the assumptions that the relatively better performance in typically-developing children is evidence that they treated the head turn as an intentional cue, and that weaker action prediction abilities in children with ASD were because they do not understand intent. If typically-developing children were also able to demonstrate similarly successful action prediction with a cue that may not be treated as intentional, then this would demonstrate that action prediction with referential gaze may simply be a result of following the direction of the head turn. Moreover, if children with ASD had difficulty with both a head turn cue as well as a cue that controlled for the direction of the head turn, then this would indicate that difficulties are not due to a lack of intention understanding, assumed from their social impairments, but a general difficulty with directional information.

In another study of action prediction, Vivanti et al. (2014) made similar conclusions when comparing children with ASD to children without ASD (a mixed group including children with global DD or typical development, n = 24 in each group). This study also included neutral and head-turning conditions where children saw the actor's actions up until the end goal, which was not shown in the video. An actor in a video spread her arms out on either side and simultaneously moved both arms down towards objects that were on either side of her hands (an object on each side). In the neutral condition, the actor faced her head to look up while moving both arms down, while in the head-turning condition, the actor turned her head down to provide referential gaze to one of the objects (i.e., target object) while moving her arms down. In contrast to the prior study, children in this study were not asked to physically act on the objects, and the authors measured children's visual attention to the actor's face, actor's action (hand), and the target object using an eye tracker.

The authors found that in children without ASD, in the head-turning relative to the neutral condition, children looked more at the target object and the actor's face, and looked less at the actor's action. The authors interpreted this pattern of visual attention as an understanding of the actor's object-directed goal in the head-turning condition, though this assumes the head-turn to be a cue indicating the actor's goals. The authors discounted that children's pattern of looking could have simply been due to a "geometrical orientation" of looking at anything in the direction of the head turn (e.g., a lean explanation), and not understanding the actor's goals of acting on the object. The authors stated that in the head-turning condition, visual attention appeared to be goal-directed because children were not simply looking at anything in the direction of gaze such as the actor's hand, but specifically at the target object. Yet without a control for the head-turn, it is unclear if this pattern of looking reflects how children without ASD would monitor their attention with any cue that similarly indicated one of the objects.

In contrast to the pattern of findings in children without ASD, visual attention in children with ASD did not differ between neutral and head-turning conditions in children's looking at the target object, the actor's face, and the actor's action. Children with ASD also looked less at the target object and the actor's face than children without ASD in the head-turning condition, whereas looking time between groups did not differ in the neutral condition. Though these findings were interpreted as a lack of appreciating the actor's goals in children with ASD, without a control cue for the head-turn, it is again unclear if children with ASD would have difficulty in the head-turning relative to the neutral condition with any cue that indicated one of the objects.

Therefore, both studies proposed that children with ASD have difficulty in predicting another's goal-directed behavior, suggesting that this is because children with ASD do not treat gaze as an intentional cue (Vivanti et al., 2011). Yet this explanation assumes that in typical development, successfully predicting the actor's goals is because children treat another's head turn and referential gaze as intentional. No studies to date have tested whether children could successfully predict another's goals by following a simple directional cue, such as an arrow, as well as they do with a head turn/referential gaze cue. It may be that both groups of children use referential gaze differently from an arrow cue, which would suggest that both groups treat referential gaze as intentional. Another possible outcome, in line with studies of Vivanti et al. (2011, 2014), is that typically-developing children treat referential gaze as intentional, whereas children with ASD as a group do not treat referential gaze differently from another directional cue, which would be indicated by similar performance between conditions with a referential gaze or an arrow cue. Finally, it could be that neither group of children would treat referential gaze differently from another directional cue of an arrow. In this study I also examined the role of intention understanding by examining if a potentially intentional reading of gaze might be seen in children's in-depth learning; this in-depth learning has not yet been tested in the context of action prediction. For example, if with referential gaze, children generalize the referent to different versions of that referent (Bani Hani et al., 2012; Csibra & Gergely, 2009), and they do not generalize when the referent was cued by an arrow.

The Current Study

This study examined whether 6- to 11-year-old children with ASD and typically-developing children use referential gaze, and potentially, another's intentions, in the context of action prediction. As noted in Chapter 1, this age range was selected to be comparable to the ages of children in prior studies of action prediction (Vivanti et al., 2011). Learning was compared between the same arrow cue and referential gaze cue as that used in the word learning context described in Chapter 1. The same sample participated in both word learning and action prediction studies at their first visit, though 2 children with ASD and 1 TD child were not included in the action prediction context (see the Chapter 2 Methods section for more information on why these children were not included). Despite removing these children, participants were still matched on nonverbal IQ, chronological age, ratio of girls to boys, and parent level of education. Again, this sample includes children with ASD who have nonverbal IQ in the normal range, representing only a subgroup of children on the autism spectrum.

Children watched a video where there was block tower (made up of two blocks) in front of an actor. Additionally, to complete the block tower, there were two additional blocks in the scene. One of the additional blocks was on the left side of the actor's head, and the other block was on the right side. Either a referential gaze or an arrow cue indicated one of the blocks, which was meant to be the third and final block that the actor would use to complete the tower. Again, all nonverbal cues (e.g., head movement, body movement) were avoided, to allow for the interpretation to be specific to referential gaze itself.

As in the word learning context, I attempted to create a discrepant learning scenario, to better ensure that children were predicting the next block by following the cue instead of following their own preference. To establish this scenario, I used block towers that created an image of a common shape using three blocks. The block tower in front of the actor was incomplete with only two blocks, and the common shape could be completed by a third final block to create a conventional (e.g., a triangle), or an unconventional shape (e.g., an uncommon shape). On one side of the actor was a distractor block, which would complete the tower to create the conventional shape (e.g., a triangle), while on the other side of the actor was the target block, which would complete the unconventional shape. Thus, children's learning was tested in a discrepant scenario, where children would potentially be initially drawn to choose the distractor block (to complete the conventional shape), but the cue would indicate the target block (to complete the unconventional shape). Children would demonstrate action prediction if they chose the target over the distractor block.

Videos were similar to the word learning context and are described in detail in the Methods section. An eye tracker recorded similar measures as seen in Chapter 1 of visual attention to the target block and the area of the cue. Children's learning, or the ability to predict the actor's goal, was also measured using pointing to the target block during the test phase of the video and in-depth learning measures. In-depth learning was assessed by whether children could generalize what they learned to a set of 2D blocks similar to those seen in the video; learning were assessed immediately after the video and one week later. Finally, learning was compared by examining how the same children pointed in the word learning and the action prediction context.

There were four key differences from my videos and the Vivanti et al. (2011) study. First, I did not include the actor's actions of building the block tower, and second, children in my study did not have the blocks at their disposable during the video. This is in contrast to the Vivanti et al. study, where children actively imitated stacking the blocks on a physical tower and then predicted the actor's unknown action. In the current study, the actor's stacking motion was removed to better parallel the videos with the word learning context in Chapter 1, where the actor does not provide any movement other than referential gaze. Third, to more specifically assess the role of referential gaze versus a head turn, as in word learning, the actor provided referential gaze without turning her head. Finally, the pattern of block building was established using conventionality rather than an alternating color sequence in Vivanti et al. (2011). These differences are returned to in the Discussion.

This study contributes to our understanding of how children with ASD predict another's actions using referential gaze and provides a first look at the specific role of the directional cue of gaze in this context relative to its potential role as an intentional cue. Furthermore, comparing the performance of individual children across contexts examines how the same children use cues across different contexts, which is important when a cue such as referential gaze is used in different scenarios in our everyday lives. These findings can help us better understand how children with ASD use referential gaze when they predict another's unknown goals.

Research Questions and Predictions

Vivanti and colleagues (2011, 2014) reported differences between children with ASD and those without ASD (i.e., typically-developing group or a mixed group of children with global DD or typical development), despite groups being well matched on language and performance IQ. Thus, I also predicted group differences on children's visual attention to the scene and/or learning performance. Based on findings in the word learning context, I predicted children would respond differently to an arrow versus a referential gaze cue. Again, I argue that differences between cue conditions are evidence of children treating gaze as an intentional cue. I predicted differences between arrow and gaze conditions on all measures, but specifically on measures where cue condition differences were found in the word learning context: attention to the cue (proportion of looking time on the cue area) and latency to the target block during the test phase.

1. **Visual attention to the target block: Do children with ASD and typically-developing children follow cues and locate the target during test? Do children attend to the target block differently between cue conditions?** I measured looking time to the target object across baseline, teaching, and test phases.
2. **Visual attention to the cue: Do children with ASD and typically-developing children attend differently to a gaze versus an arrow cue?** I measured looking time to the cue area in baseline and teaching phases and the number of looks between the cue area and the target object (i.e., contingent looks) in the teaching phase.
3. **Visual attention to the target object during test: Does latency to the target differ between gaze and arrow conditions in children with ASD and typically-developing children?** I measured the latency to first look at the target block during test (Tenenbaum et al. 2014).

4. **Learning: Do children with ASD and typically-developing children predict the actor's goal better with a gaze versus an arrow cue?** I measured correct points to the target block during test.
5. **In-depth Learning: Do children with ASD and typically-developing children demonstrate better in-depth learning immediately after the video and one week later with a gaze versus an arrow cue?** I measured selection of the target block when presented with two sets of 2D cards of the block set seen in the video, both immediately after the video and one week later. The first set was the same color as the blocks in the video and the second set was a different color.
6. **Learning across contexts: Do children with ASD and typically-developing children who learn from cues in a word learning context, also learn from cues in an action prediction context?** I compared individual children's pointing to the target object in the word learning context from Chapter 1 with their pointing to the target block in the action prediction context.

METHODS

Participants

Participants and background measures were similar as Chapter 1, however 3 children from the word learning sample were excluded from the action prediction context. Two children with ASD were excluded because of incomplete data in both cue conditions (1) or differences in study procedure (1). One typically-developing child was excluded because of experimenter error when administered the action prediction tasks (1). This resulted in 22 children with ASD and 23 TD children for the matched sample as seen in Table 18. All background characteristics were similar to that seen in Chapter 1, where groups of children significantly differed on most language measures and social communication measures, but were matched on chronological age, nonverbal IQ, ratio of girls to boys, parental education, and the number of English- and French-speaking children.

Table 18. Action Prediction: Matched Sample Characteristics

	ASD (n = 22)	TD (n = 23)	*p*	*d*	vr
Age[a]	8.93 (1.23) 6.83 – 11.33	8.75 (1.11) 6.50 – 10.50	.61	.15	1.22
Nonverbal IQ[a]	108.59 (11.97) 88 – 131	109.70 (13.51) 83 – 131	.77	-.09	.79
CELF-4 Word Classes Total[a]	9.55 (3.71) 2 – 16	12.22 (3.06) 6 – 19	.02*	-.79	1.47
CELF-4 Recalling Sentences[a]	8.09 (4.07) 1 – 14	10.96 (1.97) 7 – 15	.00**	-.90	4.29
CELF-4 Word Associations[a]	30.68 (15.02) 5 – 65	33.57 (11.34) 17 – 53	.48	-.22	1.80
Socialization subscale - Vineland[a]	76.55 (12.05) 61 – 110	109.70 (12.05) 80 – 129	.00***	-2.75	0.99
Social Communication Questionnaire[a]	21.09 (5.89) 12 – 32	4.35 (2.66) 0 – 9	.00***	3.69	4.91
Gender (M : F)	19 : 3	18 : 5	.70		
Maternal education (below : above university)[#]	12 : 12	6 : 18	.14		
Number of English and French children (En : Fr)	11 : 13	10 : 14	1		

Continuous and categorical variables were analyzed using paired sample *t*-tests and Fisher's exact tests, respectively. *p* = *p* value from significance test, *d* = Cohen's *d*, vr = variance ratio. Negative values for Cohen's *d* indicate higher values in the TD group.
[a] The values shown are the mean(*SD*) and range. * $p < .05$, two tailed. ** $p < .01$, two tailed, ***$p < .001$, two tailed.
[#]1 TD child did not provide data for mother's education so father's education was used instead.

Background Measures

Please refer to Chapter 1.

Stimuli

Block design and cue comparison information can be seen in Chapter 1.

Building blocks. Four block sets were designed for this study. Each set included four blocks with two of the blocks forming the base of a pre-stacked tower, and the other two blocks (one target and one distractor) were two different options to complete the tower. The tower would complete an image of either a conventional or an unconventional shape. The distractor block, or the block that participants would more likely choose, would result in a conventional shape (e.g., *triangle*). The target block, or the block that the cue indicated, would result in an unconventional shape. The distractor block was considered to be the block that children would more likely choose because it was more common and familiar. By having one block be the more obvious choice and potentially children's preference, this attempts to create a discrepant scenario as in the word learning study, which better ensures that children were following the cue to the target block rather than what would more likely be their preference (e.g., the distractor). Therefore, the discrepant condition established in this study would demonstrate that children followed the cue to learn that the unconventional target block completed the tower rather than following their knowledge of the conventional distractor block.

Though Vivanti et al. (2011) used an alternating color sequence, I did not use a color sequence for experimental trials in case poorer responding in children with ASD in Vivanti et al. (2011) may have been because of a potential perseveration of sameness (e.g., all three blocks in the tower needed to be the same color), sometimes noted in children with ASD (American Psychiatric Association, 2013). Details about block creation and adult pilot testing can be seen in Appendix C. Pilot testing confirmed that distractor blocks were considered to be a better way to complete the tower than the target block. There were two trials per cue condition, thus all children were taught four block sets, two sets per cue condition. Block sets were not fixed to cue conditions, and were counterbalanced such that all sets were shown with either the gaze or the arrow cue across children.

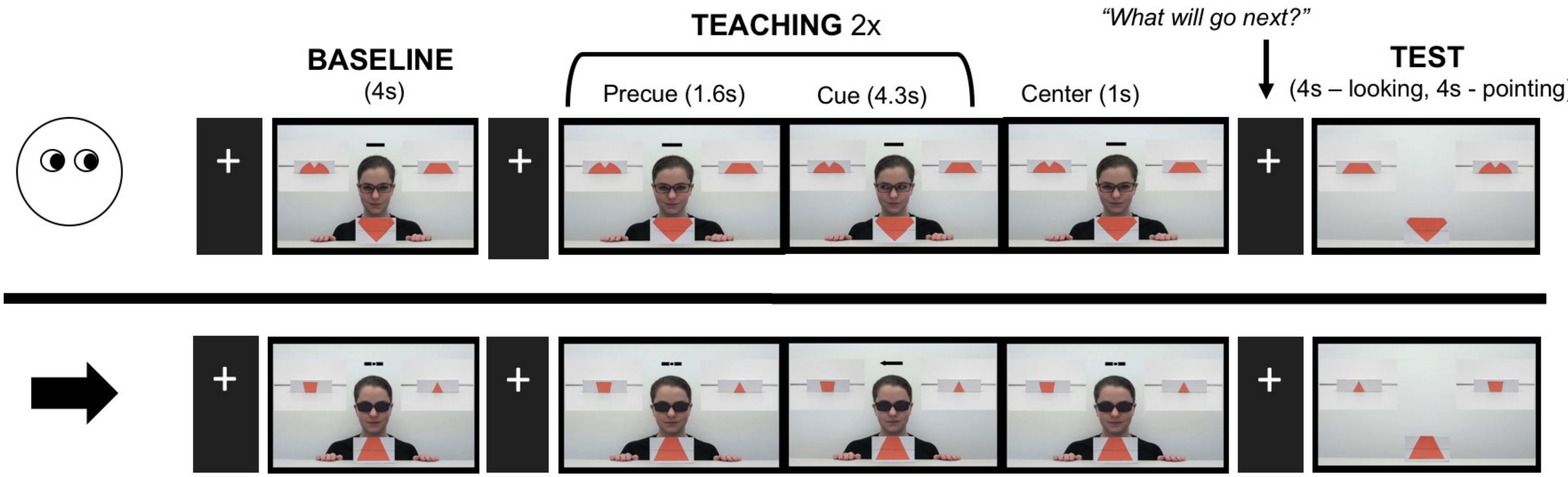
Videos. The videos of the actor were exactly the same as those used in the word learning context and can be seen in Figure 14. The distractor and target blocks replaced the word learning objects and the tower was added in front of the actor below her chin. For each cue condition, children received one practice trial and two experimental trials, for a total of two practice trials and four experimental trials. A trial refers to both the video and the in-depth learning measure presented after the video.

Whereas in the word learning study practice trials consisted of familiar objects, a parallel was not as clear for an action prediction context. However, as in the word learning context, the purpose of the practice trial in the action prediction context was to ensure children understood the procedure of watching videos, pointing to an object, and answering questions about what they saw. Therefore, I did not consider it necessary to establish practice trials with familiar objects, but to provide a different example of block building and ensure children understood when to respond. Thus, instead, I adopted a color sequence similar to Vivanti et al. (2011), though instead of alternating colors, the bottom two blocks were the same color and the choice of the top block could either be one that would continue the color sequence (distractor) or a different colored block (target).

Practice trials were administered before experimental trials in each cue condition. Children were told, "*You're going to watch a video and learn how to build a tower. I want you to pay close attention because I'm going to ask you to show me what you learn after the video.*" If children did not point to the target object during test, an experimenter provided a prompt, "*Oh* *did you notice there's something that tells you which one is the right thing? Let's watch it again!*" The goal of the practice video was not to tell children whether they were correct or incorrect, but simply to make sure children understood the procedure of watching all of the information in the video and formulating a response. If children did not point to the target (pointing instead at the distractor or did not point at all) when they were shown the video the first time, it was unclear whether they did notice the cue in the video. For this reason, the prompt would bring their attention to the video one more time, but did not reveal what to specifically pay attention to. Even if children still pointed incorrectly upon watching the video a second time, experimenters continued to the in-depth learning measure. No explicit feedback was provided throughout the trials, although experimenters did provide motivational phrases such as "*Great!*" A pass in the practice video was defined by pointing to either the target after watching the video once or even if the child still pointed to the distractor after repeated viewing of the video.

Before watching experiment videos, children were told, "*Now you're going to learn how to build other types of towers, and I want you to pay close attention because I'm going to ask you to show me what you learned after the video like we did before [with the practice video]*". The video phases were similar to those in the word learning study. All interstimulus intervals (ISI), or intervals that separated different parts of the video, were 1.5 s. **1) Baseline:** the child could view the scene for 4 s. **2) Teaching:** for ~ 1.6 s the actor provided direct gaze (referential gaze condition) or the actor was wearing glasses and there was a static circle above the actor's head (arrow condition); this portion of time is the *pre-cue portion* of the teaching phase. Next, the cue moved to indicate the target object for 3.6 s, then returned to center; from the moment the cue began to move to the moment the cue returned to center is considered the *cue portion* of the teaching phase. As in word learning, the duration of the cue portion totaled to 4.3 s, because it included the cue shift to the target, the cue indicating the target, then the cue shifting back to center. The same teaching phase was presented twice. **3) Test:** prior to seeing the objects during test, children heard the prompt, *"What will go next?"* during the ISI, then the objects were presented for 4 s. The end of this phrase was set to the last frame of the ISI, meaning children saw the test images immediately after the prompt. Target object side was switched at test to better dissociate whether children's learning was due to the cue rather than the side the cue indicated. After 4 s of viewing the objects, children had another 4 s to point to the target, "*Now point to it*" for an explicit measure of their learning, rather than looking time.

Figure 14. Action Prediction: Frames from the Video that Depict an Example of an Action Prediction Video Sequence

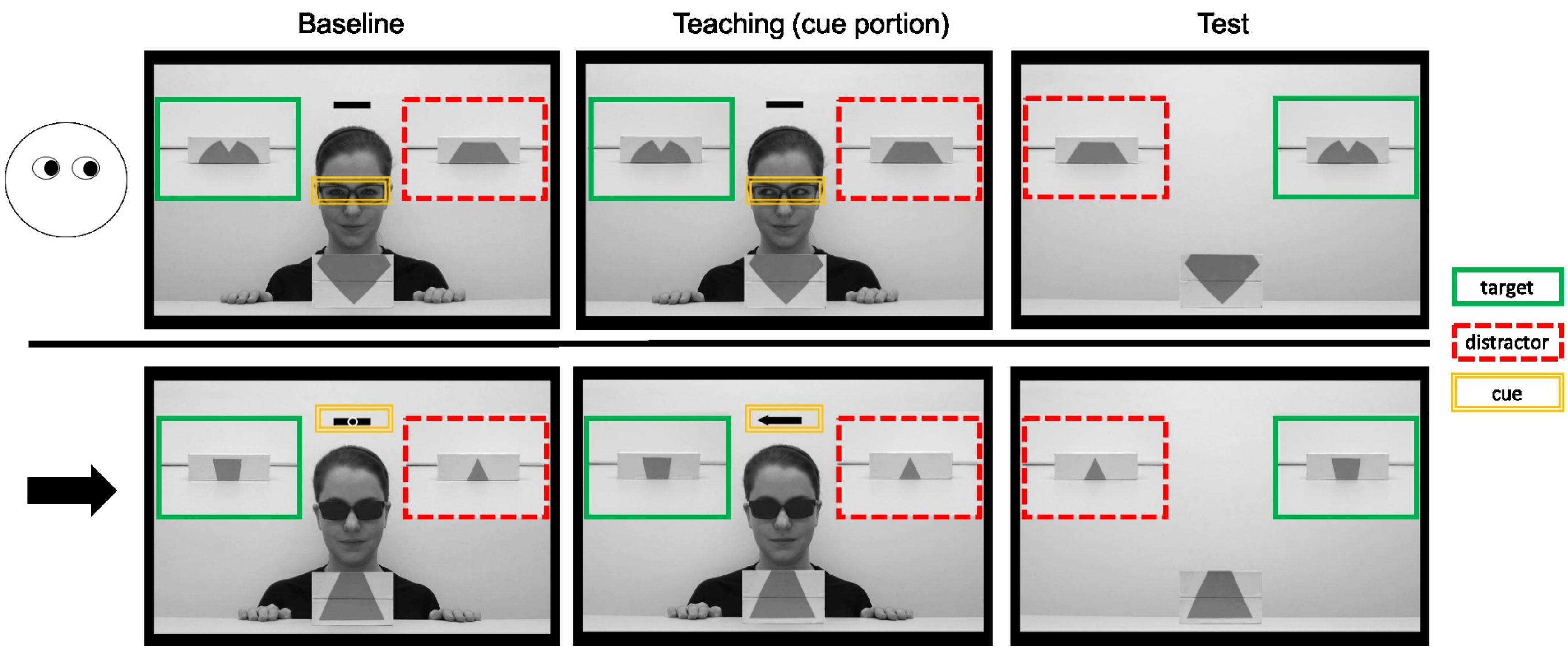


Frames from the referential gaze condition (above the solid line) and the arrow cue condition (below the solid line). The duration of the video sequence was approximately 30 seconds per block set. Children had two experimental trials per cue condition (with different block sets).

Eye Tracking Video Stimuli

The same three video phases were examined in this study as seen in Chapter 1: 1) baseline, 2) teaching, and 3) test. For analysis, data from the teaching phase included only the cue portion (see the Video section above for details), which was presented twice for each block set. The same three areas of interest were also the same size and in the same exact location as that seen in the word learning context: 1) target block (H 306 pixels and W 440 pixels) 2) distractor block (H 306 pixels and W 440 pixels) 3) cue area (not applicable at test; H 54 pixels and W 184 pixels). The video of the actress were the exact same as those used in the word learning study, and only the objects differed. See Figure 15 for the placement of the areas in the scene.

Figure 15. Action Prediction: Areas of Interest During the Baseline, Teaching, and Test Phase



To more clearly highlight the areas of interest, the video frames are shown in grayscale, though the videos were shown in color as can be seen in Figure 14. The target block is highlighted in the green solid line, the distractor block in the red dotted line, and the cue in a light orange with double lines.

In-depth Learning Measure

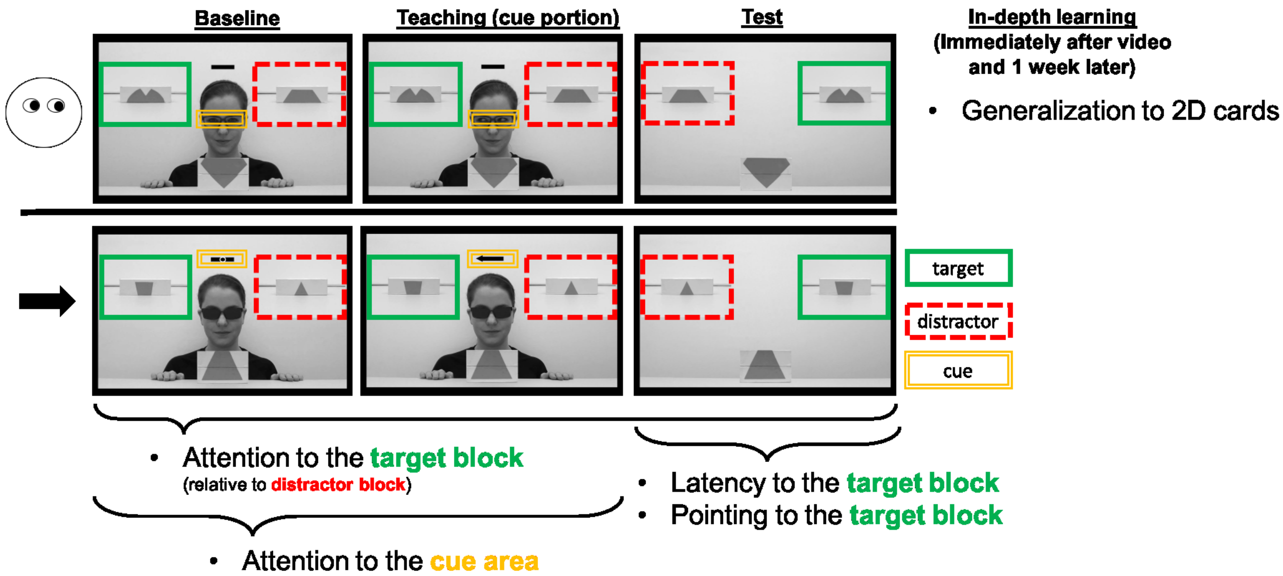
Block generalization. Children's in-depth learning of action prediction was assessed both immediately after the video and one week later in a generalization task. After each video, children were presented with two different 2D cards; an example can be seen in Figure 16. The first card had a 2D version of the same block set as seen in the video, presented with the target on the same side as seen during test in the video. Children were asked, "*Now show me what goes next*". If children wanted to manipulate the two options before choosing their final answer, they were encouraged to do so. After children completed the first card, they were given a second card where the distractor and target block were a different color than the original version. Side of the target block in the second card was counterbalanced across participants. Children's choice of the target and distractor blocks were scored as 1 and 0, respectively, for a possible score of 2 correct with the same color and 2 correct with the different color in each cue condition. An overview of all measures analyzed in this task, as well as the phases they were collected in, is presented in Figure 17.

Figure 16. Action Prediction: Example of Generalization Task

Same Color Different Color

The same color blocks on the left were the same color orange as the block seen in the video. The different color blocks on the right includes the pre-stacked tower in the same color of orange as the original set, but the target and distractor blocks were a different color than orange (brown for this shape).

Figure 17. Action Prediction: Overview of Measures Analyzed



This figure depicts when the different measures analyzed in this study were collected during testing. To more clearly highlight the areas of interest, the video frames are shown in grayscale, though the videos were shown in color as can be seen in Figure 14. The target block is highlighted in the green solid line, the distractor block in the red dotted line, and the cue in a light orange with double lines.

Measures of *attention to the target block* were collected during baseline, teaching, and test phases, *attention to the cue area* was collected during baseline and/or teaching phases, and *latency* and *pointing to the target block* was collected during the test phase. At visit 1, children were tested on their *in-depth learning* of the target block immediately after the video using the generalization to 2D cards. At visit 2, children were only tested on their in-depth learning and did not watch the video again.

Analysis Plan

Analyses were conducted similarly to Chapter 1. Continuous variables (i.e., eye tracking measures) were evaluated using linear mixed models where results focused on the matched groups of children with ASD and TD children in a 2 (cue condition: gaze versus arrow) x 2 (group: ASD vs TD) mixed design, and an additional factor was included when measures occurred over multiple time points (i.e., video phases, visits). Post hoc tests were conducted on significant effects when $p < .05$, and p values of multiple comparisons were corrected with the Tukey method and a family-wise error of alpha = .05 (Lenth, 2016). Predicted marginal means

(*PMM*s) are means adjusted for when data are unbalanced (e.g., missing data). For all transformed data, *PMM*s were back-transformed and thus can be interpreted in their original units. The difference scores between *PMM*s are similar to other difference scores between means and can be considered as unstandardized effect sizes. Standardized effect sizes are provided by Cohen's *d* (Westfall, 2016), on the means and standard deviations of the raw data. Raw data refers to data that has not been corrected for unbalanced data. All raw data are provided in tables for transparency.

For categorical variables (i.e., pointing and in-depth learning), children's level of success at learning was summarized into five levels and the distribution of levels in each group of children was compared using Fisher's exact test. These levels included: 1) *success with both cues,* meaning children identified the target in three or four trials, 2) *success with gaze* or 3) *success with arrow*, meaning children identified the target in only both trials of the respective cue condition, 4) *at chance with both cues,* meaning children only identified the target once in the gaze and once in the arrow condition, and 5) *limited or no success*, meaning children were only able to identify the target in one or none of the trials.

Eye Tracking Diagnostics

Please see Appendix G for more information on calibration and data cleaning. Groups of children did not significantly differ on their calibration metrics. Two children with ASD from the matched group were not included in eye tracking analyses because 1 child had strabismus and the other could not be calibrated. This resulted in 20 children with ASD and 23 TD children included in eye tracking analyses.

Calibration. Children were not calibrated differently for action prediction videos than for word learning videos, but because sample sizes slightly differed across contexts, children's mean angular error (measure of accuracy and precision) was compared again in the action prediction sample. Children with ASD and TD children did not significantly differ on their mean angular error, suggesting that the accuracy and precision of eye tracking data were similar between groups.

Data cleaning. The same looking time cut-off criterion of 25% looking in each phase was used as in Chapter 1. As in Chapter 1, one video sequence included baseline, teaching (pre-cue and cue portions), center, and test phases, and each child had 4 video sequences (2 per cue condition). Video sequences were considered complete for analyses when overall looking times were above 25% to baseline, cue portion of the teaching phase, and test phases. Video sequences were dropped when they were not considered complete for analyses. Six video sequences were dropped for children with ASD (affected 5 different children) and four video sequences were dropped for TD children (affected 3 different children). However, all children provided at least one complete video sequence for analyses in each cue condition.

I used a linear mixed model to examine children's overall looking time to the scene, with a dependent variable of the proportion of overall looking time. The proportion was calculated in the same manner as in Chapter 1, by dividing the overall looking time in a phase (based on a unit of 33 ms) relative to the duration of that phase; proportions ranged from 0 to 1. There were no significant differences between groups in the proportion of overall looking time to the scene. As expected, both groups spent a significantly smaller proportion of overall looking time during test relative to both earlier phases of baseline and teaching, indicating that looking time decreased over the course of the video sequence. However, in the TD group this pattern was only seen in the gaze condition, whereas in the arrow condition the proportion of overall looking time did not differ across phases. Within TD children, they spent a significantly greater proportion of overall looking time during the test phase in the arrow relative to the gaze condition. Therefore, if there were any potential differences between gaze and arrow conditions during the test phase for the following research questions, it could in part be attributed to children's overall looking time difference between cue conditions. However, for the research questions below, I found no differences between cue conditions during the test phase for TD children, thus these results are not discussed further. Please see Appendix G for more details.

RESULTS

All visual attention measures below were calculated using a unit of 100 ms fixation duration.

Visual Attention to the Target Block: Target Advantage

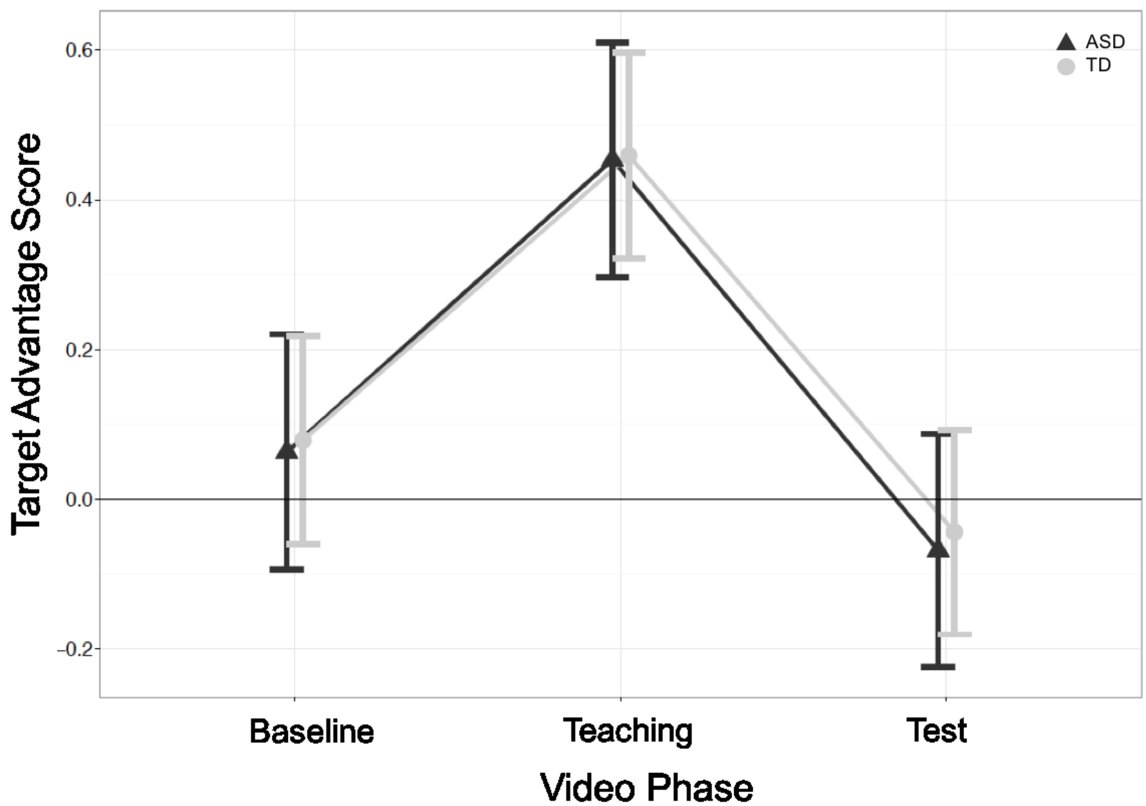
As seen in Chapter 1, children's ability to follow the cue to the target block during teaching and locate the target block during test was measured with a *target advantage score*. First, the total fixation duration to the target and distractor objects were summarized in each baseline, teaching and test phase. Then, target advantage scores were calculated by subtracting the fixation duration to the distractor (d) from the fixation duration to the target (t), and dividing by the total fixation duration to both the target and distractor, i.e., (t – d) / (t + d). The target advantage scores ranged from -1 to 1, with positive numbers indicating more looking to the target and negative numbers indicating more looking to the distractor. Target advantage scores provide a way to monitor attention to the target block across baseline, teaching, and test phases. Higher scores during the teaching phase relative to the baseline phase could indicate that children followed the cue to the target during teaching. Higher scores during test phases in comparison to baseline could indicate that children may have been influenced by the cue during teaching, thus locating the target during test. Additionally, our primary analyses examined whether looking time to the target block differed between cue conditions, and if there were any interactions of cue condition with factors of group or video phase.

The linear mixed model included a random intercept of participant and fixed effects of cue condition (gaze, arrow), group (ASD, TD), and video phase (baseline, teaching, test). There were no significant main effects of group or interactions between cue condition, group, or video phase (Fs < .42, ps > .52). The cue condition main effect was marginal, $F(1, 421.23) = 3.36$, p = .07, where all children had more of a target advantage in the arrow condition (*PMM* = .20, *SE* = .05) versus the gaze condition (*PMM* = .12, *SE* = .05). Results revealed only a significant main effect of video phase, $F(2, 410.28) = 34.07$, $p < .001$.

Post hoc tests of the video phase effect revealed that children had difficulty locating the target block during test. That is, children followed the cues to the target during teaching (*PMM* = .46, *SE* = .05) in comparison to their target advantage scores during the baseline phase (*PMM* = .07, *SE* = .05), $t(419.63) = -6.01$, $p < .001$, *PMM* difference = - .39, d = -.58. However, children did not demonstrate a target advantage during test (*PMM* = - .06, *SE* = .05) when compared to baseline, $t(419.11) = 2.00$, p = .11, *PMM* difference = .13, d = .18. The *PMM* close to 0 during the test phase indicates that children were looking similarly between the target and the distractor during the test phase, and the positive *PMM* difference between baseline and test phases indicates that children were looking more at the target during baseline than teaching phases. As expected, there was less of a target advantage during test in comparison to teaching, $t(420.02)$ = 8.07, $p < .001$, *PMM* difference = .51, d = .76.

Therefore, while watching baseline, teaching, and test phases, children with ASD and TD children followed both gaze and arrow cues to the target block during teaching, however, neither group demonstrated that they located the target block during test relative to their baseline looking. Positive target advantage scores during baseline indicated that contrary to looking more at the distractor block (which would create a conventional shape), children appeared to have an initial preference for the target block (which would create an unconventional shape). Figure 18 collapses across cue conditions and demonstrates children’s target advantage scores across phases in each group of children. Raw target advantage scores by group and cue condition can be seen in Table 19.

Figure 18. Action Prediction: Predicted Marginal Means of Target Advantage Scores Collapsed Across Cue Conditions for Each Video Phase in Each Group of Children



Points represent predicted marginal means and error bars are 95% confidence intervals. More looking to the target is shown with positive values and more looking to the distractor is shown with negative values. The horizontal line of 0 indicates equal looking to the target and the distractor.

Table 19. Action Prediction: Raw Target Advantage Scores

Baseline	ASD (n = 20)	TD (n = 23)	d
Gaze	-.01 (.59)	.01 (.59)	-.03
Arrow	.09 (.48)	.15 (.56)	-.11
d	-.18	-.24	
Teaching			
Gaze	.41 (.62)	.39 (.58)	.04
Arrow	.45 (.64)	.52 (.62)	-.12
d	-.06	-.22	
Test			
Gaze	-.09 (.50)	-.10 (.48)	.01
Arrow	-.04 (.65)	.01 (.59)	-.09
d	-.09	-.21	

The values shown are the mean(*SD*). Ranges were – 1 to 1 for all subgroups except for baseline ASD arrow condition (- .72 to 1), and test ASD gaze condition (-1 to .88). Negative Cohen's *d* values represent higher scores for typically developing children when comparing across groups and higher scores for arrow when comparing across cue conditions.

Visual Attention to the Cue: Proportion of Looking Time and Contingent Looking

Children's attention to the cue area was measured with the *proportion of looking time to the cue area* and, as well as the number of looks between the cue area and the target object, or contingent looks, which was examined in the word learning context (Norbury et al., 2010; Tenenbaum et al., 2014).

Proportion of looking time to the cue area. First, as in Chapter 1, I summarized the total fixation duration to the cue area and the total fixation duration to the scene in baseline and teaching phases. Then the proportion of looking time to the cue area was calculated by dividing the total fixation duration to the cue area by the total fixation duration to the scene. As seen in the word learning study, the proportion of looking time to the cue area was measured during both baseline and teaching phases because of the proposed relevance of direct gaze as an ostensive cue (Csibra & Gergely, 2009). Thus for the purpose of this analysis, I will refer to the area of direct and referential gaze as the area of the gaze cue and the area of the static circle and the arrow as the area of the control cue. Please refer to Figure 14 in the Methods section for a visual representation of this contrast.

The linear mixed model included a random intercept of participant and fixed effects of cue condition (gaze, arrow), group (ASD, TD), and video phase (baseline, teaching). There was some heteroscedasticity due to a large number of observations being around 0% time looking to the cue area. To try and correct for the heteroscedasticity, I first removed 6 TD children and 1 child with ASD because they never looked at the cue area (though they met data cleaning criteria of more than 25% overall looking to the scene), then transformed the proportion of looking time to the cue area using a square root transformation.

As seen in Table 20, I found a significant effect of cue condition, a significant effect of video phase, and a significant interaction of cue condition by group. The video phase effect was due to children spending a significantly greater proportion of looking time to the cue area during teaching (*PMM* = .15, *SE* = .02) than during baseline (*PMM* = .04, *SE* = .01), *PMM* difference = - .11, *d* = -.57, meaning children noticed both cues during teaching. The results of the 2-way interaction between cue condition and group differed from the word learning context. In word learning, both children with ASD and TD children looked at the cue area for a significantly greater proportion of looking time in the gaze versus the arrow condition. Yet, in action prediction, children with ASD did not differ in their proportion of looking time to the cue area in the gaze (*PMM* = .11, *SE* = .02) versus the arrow condition (*PMM* = .06, *SE* = .02), $t(246.35) = 2.06$, $p = .15$, $d = .24$, and only TD children spent a significantly greater proportion of time looking to the cue area in the gaze (*PMM* = .18, *SE* = .03) versus the arrow condition (*PMM* =

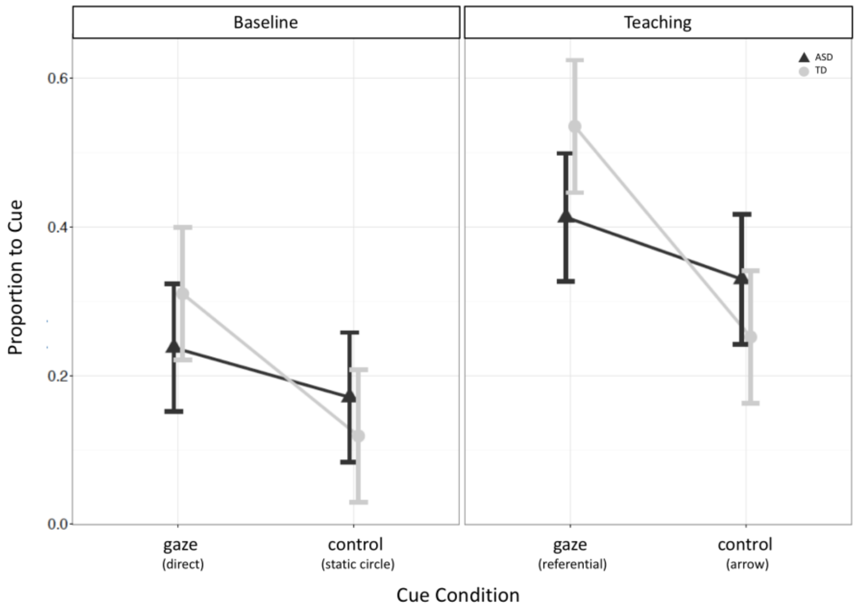
.03, *SE* = .01), *t*(242.27) = 6.42, $p < .001$, *d* = .76. There were no differences between groups (*p*s > .22). Figure 19 displays the proportion of looking time to the cue area in each phase and group. The raw data by group and cue condition can be seen in Table 21.

Table 20. Action Prediction: Main Effects and Interactions of Proportion of Looking Time to Cue Area

Factor	*df*	*F*	*p*
Cue condition (C)	1, 231.81	36.35	<.001***
Group (G)	1, 33.61	.14	.71
Video phase (V)	1, 228.28	45.20	<.001***
C x G	1, 231.81	9.80	.002**
C x V	1, 228.28	1.12	.29
G x V	1, 228.28	.05	.82
C x G x V	1, 228.28	.54	.46

** $p \leq .01$ *** $p \leq .001$

Figure 19. Action Prediction: Predicted Marginal Means of Proportion of Looking Time to Cue Area for Each Video Phase in Each Group of Children



Points represent predicted marginal means and error bars are 95% confidence intervals.

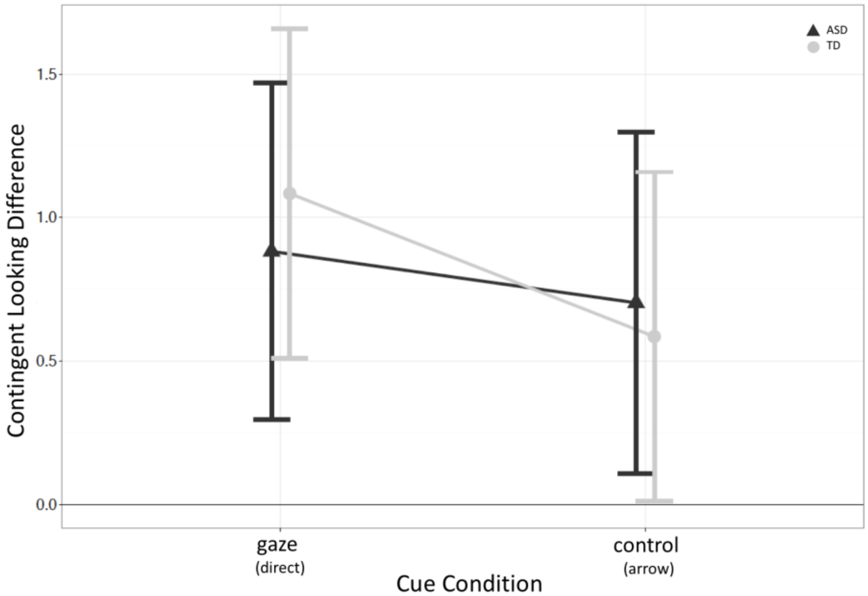
Table 21. Action Prediction: Raw Proportion of Looking Time to Cue Area

Baseline	ASD (*n* = 19)	TD (*n* = 17)	*d*
Gaze	.11 (.16)	.16 (.18)	-.30
	0 - .40	0 - .64	
Arrow	.08 (.11)	.06 (.15)	.13
	0 - .42	0 - .62	
d	.24	.61	
Teaching			
Gaze	.23 (.19)	.36 (.25)	-.56
	0 - .80	0 - 1	
Arrow	.15 (.14)	.12 (.14)	.23
	0 - .59	0 - .62	
d	.46	1.14	

The values shown are the mean(*SD*) and range. Negative Cohen's *d* values represent higher scores for typically developing children when comparing across groups and higher scores for arrow when comparing across cue conditions.

Contingent looking. As seen in word learning, I first measured the shifts between fixations to the cue area and immediate fixations to the target or distractor blocks (i.e., one contingent look), as well as shifts between fixations from the target or distractor block to the cue area. Then, I calculated a *contingent looking difference* score by subtracting the number of contingent looks to the distractor from the number of contingent looks to the target block. As seen in the word learning context, there were many instances of a difference score of 0, because children either did not provide any contingent looks or there were equal contingent looks to the target and distractor. Therefore, rather than a standardized difference score where I would not be able to divide 0 over 0, I used a contingent looking difference that was not standardized. I retained all 0 values except for 3 children (3 TD) who never provided contingent looks. Data were only analyzed during the teaching phase, when the gaze or arrow cue directed attention to the target block. The linear mixed model included a random intercept of participant and fixed effects of cue condition (gaze, arrow) and group (ASD, TD). As seen in Figure 20 there were no significant main effects or interactions between cue condition and group, (Fs < 1.72, ps $> .19$). Raw contingent looking difference scores by group and cue condition can be seen in Table 22.

Figure 20. Action Prediction: Predicted Marginal Means of the Contingent Looking Difference in Each Group of Children



Points represent predicted marginal means and error bars are 95% confidence intervals. More contingent looking to the target is shown with positive values and more contingent looking to the distractor is shown with negative values. The horizontal line of 0 indicates an equal number of contingent looking to the target and the distractor.

Table 22. Action Prediction: Raw Contingent Looking Difference Scores

Teaching	ASD (*n* = 20)	TD (*n* = 20)	*d*
Gaze	.89 (1.54)	1.08 (2.12)	-.10
	-2 to 6	-2 to 6	
Arrow	.72 (1.83)	.59 (2.17)	.07
	-4 to 5	-2 to 5	
d	.10	.23	

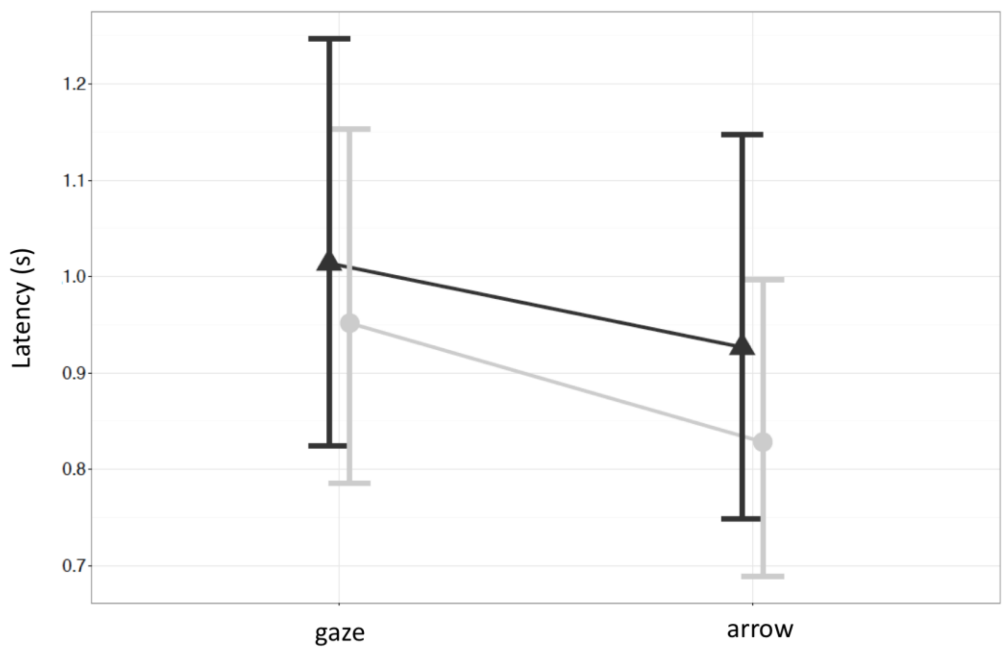
The values shown are the mean(*SD*) and range. Negative Cohen's *d* values represent higher scores for typically developing children when comparing across groups and higher scores for arrow when comparing across cue conditions.

Thus, the measures of visual attention to the cue area demonstrates that again there were no significant differences between groups, and both groups spent more time looking at the cue area during teaching relative to baseline phases. These findings suggest that both groups noticed the cues during teaching and spent a similar time looking at the cues. However, there was a difference in the relative looking between cue conditions, where TD children looked more at the area of the gaze cue relative to the area of the control cue, and children with ASD did not differ in their looking time to the cue area between cue conditions.

Visual Attention to the Target Block During Test: Latency

Although children did not demonstrate evidence of locating the target during test, I examined if their latency to fixate on the target during test differed by cue condition or group. This was motivated by differences between cue conditions on this measure in the context of word learning, as well as whether children might have first fixated to the target block, but then became unsure and thus spent similar looking to the target and distractor blocks (as seen earlier by *PMM* scores close to 0 on the target advantage score during test). Latency was measured in the same manner as seen in word learning, where it was measured from the start of the test video, which coincided with the end of the prompt heard during the ISI (i.e., "*What will go next*?"). Latencies under 200ms and over 4000ms were excluded. The linear mixed model included a random intercept of participant and the fixed effects of cue condition (gaze, arrow) and group (ASD, TD). Latencies were log transformed due to heteroscedasticity. There were no significant main effects or interactions between cue condition and group (Fs < 1.98, ps > .16). Figure 21 depicts latencies in each cue condition and group. Raw data by cue condition and group can be seen in Table 23. Therefore, in contrast to results seen in the word learning context, there were no cue condition differences in the action prediction context on the measure of latency.

Figure 21. Action Prediction: Predicted Marginal Means of Latency to the Target Block During Test in Each Group of Children



Points represent predicted marginal means and error bars are 95% confidence intervals.

Table 23. Action Prediction: Raw Latency Scores to the Target During Test (ms)

	ASD (*n* = 20)	TD (*n* = 23)	*d*
Gaze	1.14 (.57)	1.11 (.71)	.03
	.27 – 3.16	.26 – 3.16	
Arrow	1.03 (.53)	.95 (.47)	.16
	.39 – 2.72	.30 – 3.04	
d	.19	.27	

The values shown are the mean(*SD*) and range. Positive Cohen's *d* values represent lower scores (faster) for typically developing children when comparing across groups and lower scores (faster) for the arrow when comparing across cue conditions.

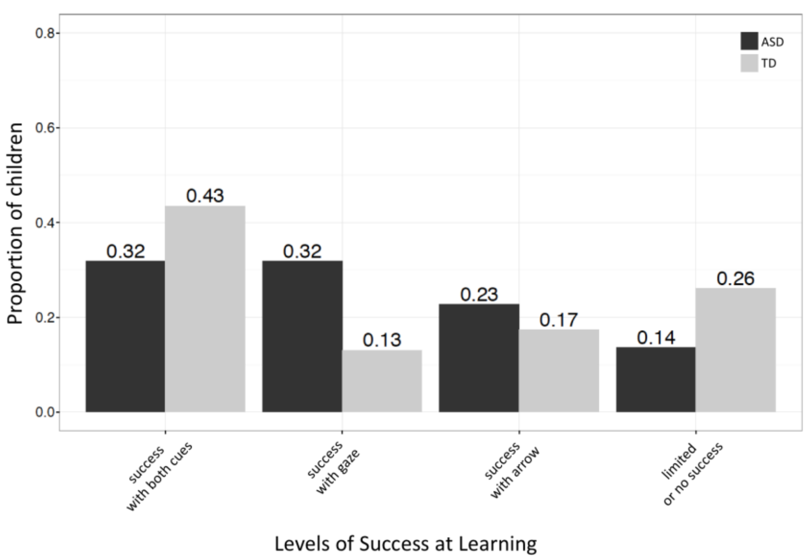
Learning: Action Prediction

At the end of the test phase, children were asked to point to the target block for a measure of their learning, or action prediction abilities. I provide results of children's pointing to the target block during the practice block set and the experimental block sets below. Children's action prediction abilities were summarized into levels of success at learning, and the distribution of the levels in each group was compared inferentially with Fisher's exact test.

Practice block sets. Practice sets used the color sequence blocks similar to that of Vivanti et al. (2011). During the practice sets, children had the opportunity to re-watch the videos if they did not point to the target or did not point to either block. If children pointed to the distractor even after watching the video a second time, they did not re-watch the video for a third time. Two participants (2 TD) watched the video three times, because they still did not understand when to point after the second time they watched the video. The same number of children with ASD and TD children needed prompts (17 ASD, 17 TD). The number of prompts required did not differ by cue condition (gaze 14 ASD, 14 TD; arrow 11 ASD, 12 TD). Less than half of the children in each group switched their answers, with only 7 children with ASD and 7 TD children switching their pointing responses after re-watching the practice video. Because there were few children who switched their pointing response, only pointing results after prompting is discussed here. Practice trials included all 22 children with ASD and 23 TD children in the matched group.

There was one practice trial before each cue condition, thus there were only two practice trials total. This reduces the levels of success at learning to four levels instead of the five levels seen for experimental measures: 1) *success with both cues* (2 correct points), 2) *success with gaze* (1 correct point), 3) *success with arrow* (1 correct point), and 4) *limited or no success* (2 incorrect points). Though the possibility of 1 correct point in the level of success with gaze or the arrow cue is not a strong indicator of success, I included these levels to maintain consistency with the distribution of levels in the experimental block sets. Figure 22 depicts the proportion of children in each level due to slightly unequal sample sizes between groups. After prompting, there were no significant differences between groups in the distribution of their levels of success at learning, Fisher's exact test $p = .39$. There were slightly more TD children that had success with both cues (10 TD) than children with ASD (7 ASD). In both groups, children who did not have success with both cues appeared to be split evenly across the other three levels: success with gaze (3 TD, 7 ASD), success with arrow (4 TD, 5 ASD), limited or no success (6 TD, 3 ASD). The pattern of responding suggests that on practice block sets, while some children in both groups appeared to have success with both cues (2 trials), over half of the children in both groups only pointed correctly once or did not point correctly on either trial.

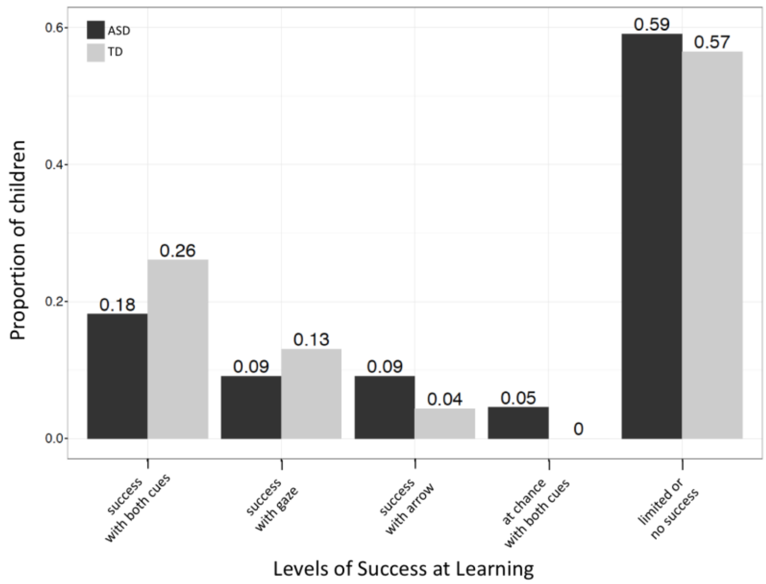
Figure 22. Action Prediction: Children's Level of Success at Learning for Pointing (After Prompt) for Practice Trials



This figure depicts the proportion of children that were in each level of success at learning on the measure of their pointing during the test phase of the practice block sets.

Experimental block sets. In contrast to practice block sets, with experimental block sets, I examined whether children would choose a target block that would complete an unconventional shape over a distractor block that would complete a conventional shape. Levels of success at learning was summarized into all five levels, because there were two possible trials for each cue condition which allowed for the level of being *at chance with both cues*. There were no significant differences between groups in their distribution of levels, Fisher's exact test p = .87. In contrast to the word learning context where the majority of children in both groups had success with both cues on the measure of pointing, in the action prediction context, both groups of children had difficulty. As depicted in Figure 23, over half of the children in both groups had limited or no success (13 ASD, 13 TD). Few children were at chance with both cues (1 ASD, 0 TD), or had success with only gaze (2 ASD, 3 TD) or the arrow (2 ASD, 1 TD). In fact, only 4 children with ASD and 6 TD children had success with both cues on this measure of pointing to the target block during test, meaning that they chose the target block on three or four trials. These findings indicate that both groups of children demonstrated difficulty with action prediction, as measured by their pointing abilities.

Figure 23. Action Prediction: Children's Level of Success at Learning for Pointing to the Target Block



This figure depicts the proportion of children that were in each level of success at learning on the measure of their pointing during the test phase of experimental block sets.

In-depth Learning Immediately After the Video and One Week Later

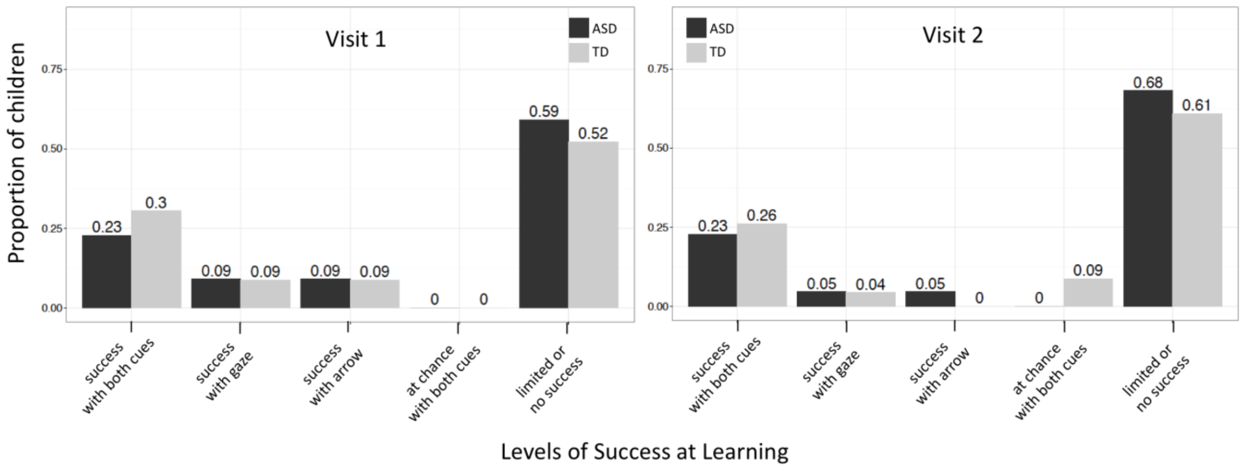
Immediately after the video and one week later, children were provided 2D cards that presented the block set that they saw in the video. The first set included blocks that were the same color as those seen in the video, and the second set included blocks where only the target and distractor blocks were a different color from those seen in the video. Children were asked to select the next block that they thought completed the tower, for a measure of whether they generalized what they learned in the video to a different set of blocks that were two dimensional, as well as a different color. I provide results of children's generalizations during the practice block sets and the experimental block sets below. Children's action prediction abilities were summarized into levels of success at learning, and the distribution of the levels in each group was compared inferentially with Fisher's exact test. All 22 children with ASD and 23 TD children were included.

Practice block sets. Children were shown a 2D version of the same blocks seen in the video. A different color version was not shown for the practice block sets. As seen in pointing, practice trials included only four levels of success at learning: 1) success with both cues, 2) success with gaze, 3) success with arrow, and 4) limited or no success. At visit 1, there were no significant differences between groups, Fishers exact test $p = .88$. Children with ASD and TD children had similar distributions of levels of success at learning: success with both cues (7 ASD, 8 TD), success with gaze (5 ASD, 4 TD), success with arrow (2 ASD, 4 TD), and limited or no success (8 ASD, 7 TD). The pattern of responding at visit 1 indicates that more children appeared to be responding at random (success with gaze/arrow since this included only one correct trial) or had difficulty with the task rather than success with both cues. However, at visit 2, there was a significant group difference in the distribution of levels, Fisher's exact test $p = .04$. There were more TD children than children with ASD who had success with both cues (2 ASD, 8 TD), and 2 children with ASD who had success with the arrow, although the majority of children in both groups had limited or no success with either cue (18 ASD, 15 TD). Thus, one week later, though the majority of children in both groups were unable to choose the target block in the practice block set (meaning that instead they chose the distractor block), there did appear to be a group difference in the distribution of children. This difference appears to be due to more TD children than children with ASD who had success at choosing the target block with both cues.

Experimental block sets. Levels of success at learning were summarized into five levels. For each experimental block set, children were presented two possible chances to generalize learning: to generalize what they learned to a 2D set that were the same color blocks that they saw in the video and a second 2D set where the target and distractor blocks were a different color from the blocks seen in the video. However, most of the children in both groups were consistent in their selections from the same color set to the different color set. For example, at visit 1, only 2 children with ASD and 2 TD children changed responses when comparing generalizations. Therefore, for simplicity, I only presented data on children's generalizations with the different color set. This is because generalizing to the different color set would be potentially more indicative of generalizing learning since the same color set could be perceived as the same set as that seen in the video and not a generalization.

There were no significant differences between groups in the distribution of their levels of success at learning at visit 1, Fisher's exact test $p = .96$ or visit 2, Fisher's exact test $p = .80$. As seen in the pointing data, many children had difficulty with learning that the target block completed the tower to create an unconventional shape. Figure 24 depicts the proportion of children in both groups for the different levels. At visit 1, few children in both groups had success with both cues, meaning that they chose the target block on three or four trials (5 ASD, 7 TD), and few had success with only gaze (2 ASD, 2 TD) or the arrow cue (2 ASD, 2 TD). Instead, around half of children had limited or no success with either cue (13 ASD, 12 TD). The distribution of levels was similar at visit 2, with over half of children having limited or no success with either cue: success with both cues (5 ASD, 6 TD), success with gaze (1 ASD, 1 TD), success with arrow (1 ASD, 0 TD), at chance with both cues (0 ASD, 2 TD), and limited or no success with either cue (15 ASD, 14 TD). Thus, many children in both groups had difficultly generalizing learning to a different color set at either visit.

Figure 24. Action Understanding: Children's Level of Success at Learning for Generalization to a Different Color 2D Block Set



This figure depicts the proportion of children that were in each level of success at learning on the measure of children's generalization to the 2D block set where the target and distractor objects were a different color from the original blocks seen in the video.

In summary, across all the potential differences on action prediction measures, the only differences found between cue conditions was in the proportion of looking time to the area of the cue, where only TD children spent more time looking at the area of the cue in the gaze relative to the arrow condition, while this cue condition difference was not significant in children with ASD. Moreover, on the practice block sets, there was one significant group difference one week later, which was seen on the measure of generalizing learning to a 2D set of the same color practice blocks. This difference suggests that while the majority of children in both groups had limited or no success on this measure, in the second visit there were four times more TD children who chose the target block than children with ASD (8 TD, 2 ASD), as well as 2 children with ASD and no TD children who had success with the arrow (though the success with the arrow should be considered with caution because this reflects only one correct trial). A summary table of all action prediction results can be seen in Table 24.

Table 24. Action Prediction: Summary of Results

Type of Measure	Measure	Difference between gaze versus arrow?
Visual Attention	Attention to the target block	✗
	Attention to the cue: Proportion looking time to cue area	✓ (TD only)
	Attention to the cue: Contingent looking to target block	✗
	Test: Latency to target block	✗
Learning	Test: Action Prediction (pointing) *Note: floor effects*	✗
In- depth learning	Generalization to 2D cards *Note: floor effects*	✗

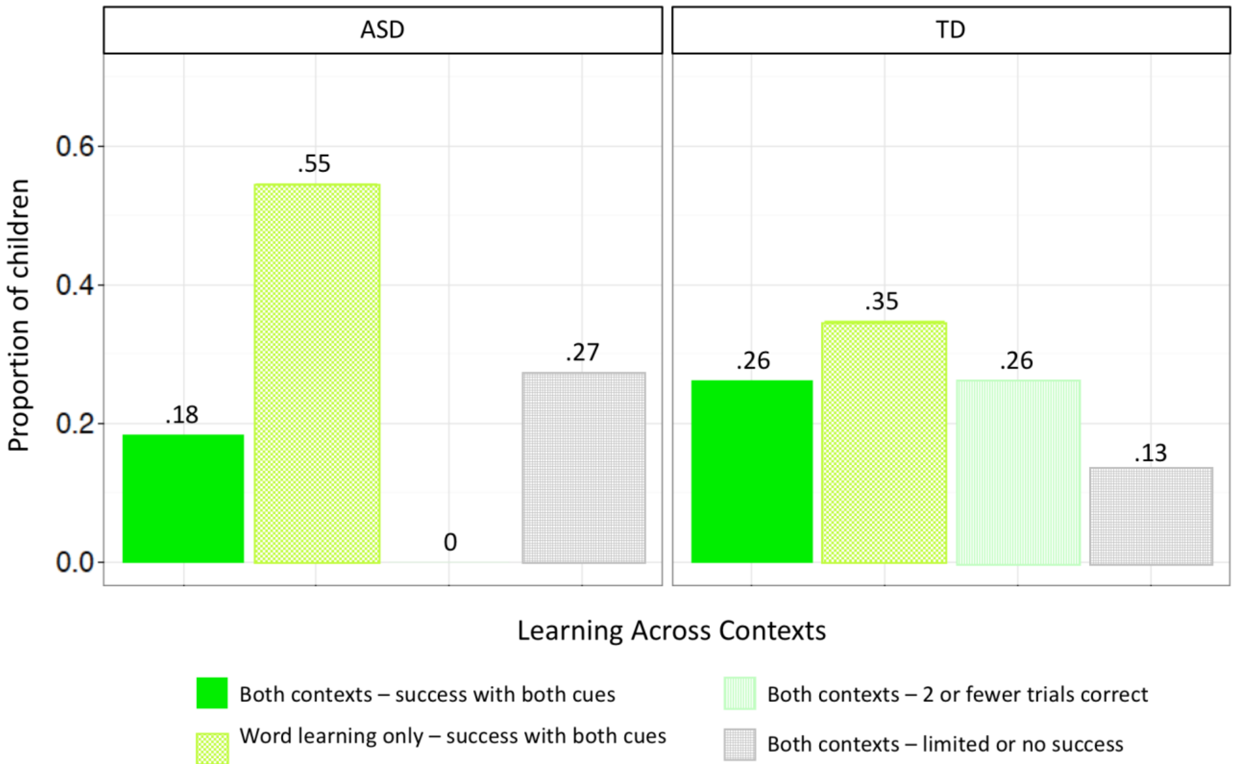
This figure depicts a summary of the findings regarding cue condition differences. ✗ = no significant differences between gaze and arrow cues (shown in black), ✓ = significant differences between gaze and arrow cues (shown in green). ✗ = floor effects (pointing, generalization), suggesting caution when interpreting these non-differences (shown in a light grey).

Pointing Across Contexts of Word Learning and Action Prediction

Finally, I examined how individual children learned from cues in the word learning and the action prediction context. I used the measure of pointing, because this was collected using a similar protocol in both contexts. Because there were so few cue condition differences noted in either context, I did not focus on cue condition differences. Instead, I examined how individual children learned from one context to another. For example, I examined if a child had success with both cues in the word learning context and whether that same child had success with both cues in the action prediction context. By examining how children use cues to help them learn across contexts, when contexts are similar in how cues and objects are presented, I can see if despite differences in what children are learning (i.e., word recognition, action prediction), they can use cues similarly or differently across contexts.

I included all children with pointing data from the action prediction context, 22 children with ASD and 23 TD children, since this was a subset of the word learning sample. Children in each group were subdivided into four levels that summarized their performance across contexts: 1) *success with both cues in both contexts* (identified the target on three or all four trials in both contexts), 2) *success with both cues in the word learning context only* (in the action prediction context these children identified the target on two or fewer trials), 3) *success in two or fewer trials in both contexts*, and 4) *limited or no success in both contexts* (identified the target only once or never in both contexts). The distribution of children in the four levels was compared between groups. There was a significant difference between groups in the distribution of these levels, Fisher's exact test $p = .04$. Figure 25 below depicts this distribution using proportions, due to the slightly uneven sample sizes in each group of children.

Figure 25. Children's Learning for Pointing Across Contexts of Word Learning and Action Prediction



This figure depicts the proportion of children that were in each level of success for how children learned across both contexts.

The levels of *success with both cues in both contexts*, and *success in two or fewer trials in both contexts*, demonstrated those children who were able to use cues similarly in both contexts to help them learn, while *limited or no success in both contexts* includes children who treat both cues similarly though this is because they are not able to use these cues to help them learn. A similar number of children in both groups had success with both cues in both contexts (4 ASD, 6 TD), though more TD children had success in two or fewer trials in both contexts (0 ASD, 6 TD). More children with ASD were those that had limited or no success in either context (6 ASD, 3 TD). Finally, there were more children with ASD than TD children who only had success in the word learning context (12 ASD, 8 TD).

These findings indicate that while both groups appeared to have more success in the word leaning context relative to the action prediction context, there was a difference in the distribution of levels between groups. This difference appears to be in part due to more TD children, relative to children with ASD, using cues similarly across contexts at different levels of learning. No children had success with both cues in only the action prediction context and not in the word learning context.

DISCUSSION

This study tested how children used referential gaze to predict an actor's unknown goal, by comparing whether children treated referential gaze as a directional or intentional cue. The action prediction context is potentially a strong test of intention understanding because successfully predicting the actor's unknown goal on an object may represent children reasoning of what the person intends to do with an object. In this context, I compared whether children followed a simple directional cue, an arrow, in contrast to a referential gaze cue, considered to reflect another's intentions. The cues used in this study were the same as in the word learning context presented in Chapter 1, thus were matched on physical features (i.e., size, color contrast, duration and angle of motion, relative distance from the target objects), and the presence of the actor. Thus, any differences between cue conditions could be explained by the proposed intentional reading of referential gaze. I compared children's performance between cue conditions using measures of visual attention to the target and the area of the cue, ability to predict the actor's unknown goal, as well as introducing a new measure, children's generalizations of the actor's goal. Children's generalizations were examined immediately after the video and one week later.

In contrast to the word learning context, I found only one cue condition difference. Specifically, typically-developing children spent a greater proportion of looking time to the cue area in the gaze relative to the arrow condition, and I did not find this cue condition difference in children with ASD. At first glance, this difference could be interpreted as a discrepancy between groups, in that typically-developing children used an intentional reading of gaze in the action prediction context and children with ASD did not. However, this group difference was not seen in children's pointing or generalization abilities, as less than half of children in both groups had success with both or one cue in either context. These findings demonstrate that while children could use both cues to help them learn in the word learning context, using these cues was more difficult in the action prediction context for children with ASD and typically-developing children. Possible reasons for this difficulty could be because action prediction is a more difficult task in general, or because of the methodological choices in the design of this study.

In this discussion I first review why children in both groups may have had difficulty with this task and then I consider what we can uncover about the intentional understanding of gaze by comparing findings across contexts.

Difficulties with the Action Prediction Task

As seen in children's pointing abilities, action prediction was similarly difficult for both groups of children. Only 18% of children with ASD and 26% of typically-developing children had success with both cues (i.e., choosing the target block on three or all four trials), while the majority of the remaining children had limited or no success with either cue (59% ASD, 57% TD). The lack of successful pointing was also replicated in children's generalization abilities to the different color set of 2D cards at both visits. At visit 1, only 23% of children with ASD and 30% of typically-developing children had success with both cues, while 59% of children with ASD and 52% of typically-developing children had limited or no success; the number of children who had limited or no success increased one week later (68% ASD, 61% TD). Though action prediction tasks have been shown to be difficult for children with ASD relative to children without ASD (Vivanti et al., 2011; 2014), the current findings demonstrate that the majority of children in both groups could not successfully predict the target block to build the tower in this study.

One reason that the action prediction context may be difficult is that successful action prediction may require more of children than in the word learning context. In the word learning context, children heard a label and followed a cue, thus learning a contingent association between a label and a target object, and then children were asked to recall the same contingent association during learning measures. Learning this contingent association may have been more concrete than in the action prediction context, where children only followed a cue to a block, but learning meant that they had to demonstrate something they hadn't been explicitly taught (e.g., which block goes on the tower). Therefore, it may be that in the action prediction context, children needed to go beyond understanding that the actor is looking at an object, and instead reason that the actor intends to use the object in a specific way. This level of reasoning may make learning in this context more difficult than in the word learning context.

Additionally, learning in the action prediction context may have also been difficult because of the methodological choices made when designing this study. One choice was to use conventional and unconventional shapes for experimental trials, rather than an alternating color pattern used by Vivanti et al. (2011). I chose different stimuli under the hypothesis that weaker performance of children with ASD in Vivanti et al. (2011) may have been due to a perseveration on the same color pattern (American Psychiatric Association, 2013), thus interfering/competing with the use of referential gaze. However, in my experiment, I found that children in both groups perseverated on the distractor block (creating a conventional shape) during test rather than choosing the target block (creating an unconventional shape), despite even initial baseline preferences to the target block and more looking at the target block during teaching. Therefore, the difficulty in this task for all children may have been in part due to an inability to break their knowledge of conventional, familiar shapes. Because children did not appear to visually locate the target during test, it is not surprising that this later translated to difficulties in pointing to the target shape and generalizing learning to target-like shapes.

Another methodological choice that may have added difficulty was the limited use of other body cues in this task relative to prior studies of action prediction (Meltzoff, 1995; Parish-Morris et al., 2007; Vivanti et al., 2011). Because this study focused on the specific role of referential gaze, I limited all other movements that were previously included in action prediction studies such as a head turn. Moreover, I did not include the action of stacking by the actor. This stacking action was removed to facilitate comparisons with the word learning context, which did not include movement other than referential gaze. Because the stacking action was removed from the video, children did not need to re-enact or imitate the actor's movements, which was also a feature of prior studies that were interested in children's imitation, as well as their action prediction abilities. Therefore, the videos in the current study provided minimal information regarding the purpose and goal of the task, though the purpose and goal were provided verbally in the prompt, "*Now you're going to learn how to build other types of towers, and I want you to pay close attention because I'm going to ask you to show me what you learned after the video like we did before [with the practice video]*". Nevertheless, both children with ASD and those with typical development may have had difficulty with predicting the actor's goal-oriented behavior because the goal was not clearly discernible, limiting children's understanding of the task itself.

Given that both children with ASD or typical development were similar in their lack of predicting the actor's unknown goal, it is difficult to interpret the cue condition difference found in visual attention for typically-developing children. On the one hand, it may be that there is a difference between how children with ASD and typically-developing use referential gaze in the action prediction context, whereas typically-developing children treat gaze as intentional and children with ASD simply use gaze to direct their attention. Therefore, this data would provide evidence to support a rich interpretation of gaze in the action prediction context (Parish-Morris et al., 2007; Vivanti et al., 2011), albeit referential gaze alone may not be a strong enough cue to influence learning in this study. However, children with ASD as a group also spent a greater proportion of looking time to the cue area in the gaze relative to the arrow condition, though this effect was smaller and not significant (Cohen's *d*s: ASD = .24, TD = .76; *p* values: ASD = .15, TD < .001). These findings suggest individual differences among children with ASD in the proportion of looking time to the cue area, which could reflect individual differences in children who do or do not treat referential gaze as an intentional cue in this context. Nevertheless, the lack of successful learning in either group complicates the interpretation of gaze as an intentional cue, since this potential intentional reading of gaze did not appear to be used in children's action prediction abilities. The lack of between group differences on visual attention measures is contrary to prior action prediction studies who found group differences despite matching children on similar variables of chronological age, IQ and even language abilities (Vivanti et al., 2011; 2014). Though some differences with prior studies were noted above regarding the stimuli and limited body cues, another difference that may have affected children's visual attention to the scene was because the area of the cue was restricted to the area around the cue specifically (i.e., referential gaze) in contrast to prior studies that used the area around the actor's face. Thus, when the area of looking is restricted to the area around gaze, there may be fewer group differences as seen in Chapter 1.

There was one significant group difference in the distribution of levels of success at learning. This difference was found on the practice set of blocks (using a color sequence) on the measure of generalization to 2D cards, but only one week later. At the second visit one week later, there were more typically-developing children (2 ASD, 8 TD) who remembered to use the target block with both cues (a different color from the rest of the tower), rather than the distractor block (the same color as the rest of the tower), although most children in both groups chose the distractor block, demonstrating limited or no success with either cue (18 ASD, 15 TD). These findings are similar to those in Vivanti et al. (2011) who also used a color pattern (alternating color sequence) and found better action prediction immediately after learning in typically-developing children relative to children with ASD. It may be that a color pattern is easier than a shape pattern, particularly for typically-developing children. However, it is important to note that for these few children, learning is seen with both cues, thus does not speak to intention understanding, and the task was still difficult for the majority of children in both groups.

Therefore, in the action prediction context, children in both groups were unable to use referential gaze or an arrow cue alone to predict the actor's next action, as the majority of children did not demonstrate successful action prediction with either cue on children's pointing or in-depth generalizations to 2D cards. However, the finding of increased looking to the area of gaze relative to the area of the control cue provides some evidence that could suggest that referential gaze is read as an intentional cue in typically-developing children in the action prediction context. This rich interpretation would be in line with others who have assumed intention understanding in typically-developing children, but did not directly test a lean versus a rich interpretation of gaze (Vivanti et al., 2011; 2014). Moreover, this direct test between an arrow cue that controls for the direction of attention and a potentially intentional referential gaze cue allows for a more nuanced interpretation of findings in children with ASD. In other words, though as a group, children with ASD did not look significantly more at the area of the gaze cue relative to area of the control cue, this was still a positive effect. This positive effect suggests that if more looking to gaze relative to a control cue is indicative of intention understanding, then there may be individual differences in children with ASD who do or do not treat referential gaze as intentional in the action prediction context. Future studies that include clearer stimuli and action movements should continue to explore stringent tests of a referential gaze versus a control cue to better understand the role of intentionality in this context.

Referential Gaze Across Word Learning and Action Prediction

Though intention understanding was difficult to determine in the action prediction context on its own, another way to assess the role of intention understanding during action prediction is by comparing findings across contexts. Because intention understanding was established in the word learning context, if children are using referential gaze consistently across contexts, then this would suggest that intention understanding is important in both word learning and action prediction. Below I discuss both global differences in children's visual attention across contexts and review how the same children demonstrated learning, using the measure of pointing, in both contexts of word learning and action prediction.

The progression of children's visual attention to the target object differed across contexts. In the word learning context, children initially attended to a distractor (more perceptually salient object), followed the cue to the target (less salient object), and then identified the target during test. In the action prediction context, children attended to the target first (that would create an unconventional shape), followed the cue to the target, and then identified the distractor (that would create a conventional shape) during test. In other words, whereas in the word learning context children initially attended to the distractor, but identified the target object during test, in the action prediction context, children initially attended to the target block, but did not choose the target block during test. Moreover, latency findings differed across contexts. Whereas in the word learning context children were significantly faster at locating the target in the gaze relative to the arrow condition, in the action prediction context there was neither a significant cue condition difference nor the same direction of the effect, with instead slower latencies in the gaze relative to the arrow condition. Thus these differences across contexts in children's attention to the stimuli, demonstrates that the stimuli in the contexts were not attended to in the same way.

Notwithstanding the differences in visual attention to the target, there were some similarities across contexts in visual attention to the area of the cue. First, across both contexts, children in both groups spent a greater proportion of time looking at the area of the gaze relative to the area of the control cue, though this was not significant in children with ASD in the action prediction context. This longer looking to the area of the cue in the gaze relative to the arrow condition was considered as one key finding of intention understanding in the context of word learning (in combination with findings in latency and semantic features). Second, though there were no significant cue condition differences in children's contingent looking difference in either context, children in both groups demonstrated a slight preference of more contingent looking to the target (relative to the distractor) in the gaze relative to the arrow condition. As noted, these cue condition differences were not significant, but were always positive effect sizes in each group of children across contexts (proportion of looking time to the cue area $d = .42 – 1.14$; contingent looking $d = .10 – .38$). Hence, visual attention to the area of the cue indicates a similar preference for gaze in both groups of children and in both contexts.

In sum, though there was a preference for gaze in attention to the area of the cue across contexts, this preference was not seen in children's visual attention to the stimuli, particularly during test. For example, whereas latencies during test in the word learning context were faster with gaze, this was the opposite case in the action prediction context, with slower latencies to locate the target in the gaze relative to the arrow condition. One reason for these differences may be in the stimuli used across contexts, since stimuli in the word learning context were made using perceptual salience and stimuli in the action prediction context were made using conventional salience. Nevertheless, the similarities across contexts in the preference to the area of the cue in the gaze relative to arrow condition provide some preliminary evidence that children may treat referential gaze as intentional in both contexts, though this may be more heterogeneous in children ASD in the action prediction context.

Finally, to examine if action prediction was generally more difficult for children, I examined how individual children used both referential gaze and arrow cues to help them point at the target in both word learning and action prediction contexts. There was a significant group difference in the distribution of levels of success across contexts. Whereas typically-developing children were more spread out, including the full range of children who had success with both cues in both contexts to those who had limited or no success in either context, children with ASD were divided into one less group. It appeared that fewer typically-developing children than children with ASD had success only in the word learning condition (18% ASD, 26% TD), but more typically-developing children had success in two or fewer trials in both contexts (0% ASD, 26% TD). Notably, there were children in both groups had success with both cues in both contexts (18% ASD, 26% TD) and children who had limited or no success in both contexts (27% ASD, 13% TD). Overall, these findings indicate that more typically-developing children used cues similarly across both contexts, though this was spread out across different levels of success in both contexts. Yet both groups of children had more success in the word learning than the action prediction context, and no children had success with both cues in the action prediction context and not the word learning context, which suggests that in the presentation of the contexts in this study, it was more difficult to use cues in the action prediction context.

Conclusions

The results in the action prediction context exhibited that children with ASD and typically-developing children were unable to follow a referential gaze cue or a directional arrow cue to predict the actor's goal. That is, children in both groups more often followed their own knowledge of conventional shapes (e.g., a triangle) rather than choosing a block that would create a new, unfamiliar shape. This demonstrates that following any cue to help children learn in the action prediction context is much more difficult than in the word learning context. Though there was some indication that typically-developing children treated referential gaze as intentional, by looking more at the area of gaze relative to a control cue, this gaze advantage was not seen in their action prediction or generalization abilities. If typically-developing children in this task do treat referential gaze as an intentional cue, this would support others who suggest that task success requires intention understanding of another's gaze direction (Vivanti et al., 2011; 2014). However, rather than a deficit in children with ASD, there may be individual differences in treating gaze as an intentional cue. Future studies must address how children attend to a referential gaze cue relative to a directional cue when children can demonstrate successful action prediction. For example, setting up a clearer context of action prediction by including the actor's movements or including stimuli other than block building (e.g., putting a pen in a cup; Leighton, Bird, & Heyes, 2010). These future directions will provide more conclusive insight into the role of intention understanding in this context.

GENERAL DISCUSSION

The intentional meaning of referential gaze is important to the fundamental understanding of this cue in social communication and child development. However, this intentional meaning also has clinical importance for children with ASD, if their social impairments result in making it difficult for them to treat referential gaze as an intentional cue that may help them learn about the world. The goal of this dissertation was to examine whether children with ASD and typically-developing children learn using a directional or potentially intentional reading of referential gaze. This was tested in two contexts of learning: word learning and action prediction. Moreover, I investigated whether children's in-depth learning, beyond initial word recognition or action prediction abilities, differed between a referential gaze or an arrow cue. This in-depth learning was examined both immediately after the video and one week later. Better in-depth learning with referential gaze relative to an arrow cue would reflect important consequences of an intentional reading of gaze.

Prior studies have examined if children can demonstrate in-depth word learning (Norbury et al., 2010), use referential gaze to predict another's actions (Vivanti et al., 2011; 2014), or learn across contexts (Parish-Morris et al., 2007). However, these studies have all interpreted findings by assuming intention understanding without ruling out alternative lean explanations for how referential gaze could be used to contribute to learning. Though Field (2016) did examine how children can learn new words with cues other than referential gaze (i.e., an arrow cue, an illumination cue), the cues were not well matched, making it difficult to compare learning with different cues. This dissertation was novel in bringing together these studies to test the assumption of an intentional reading of referential gaze. This assumption was tested by comparing how children with ASD (within the normal range of IQ) or with typical development learn with a potentially intentional cue of referential gaze versus the directional control of an arrow cue that was matched on size, motion, and physical features, such that differences between cues may be due to an intentional reading of gaze. By using this within-group comparison, I provided evidence that a referential gaze cue may be treated as an intentional cue in both children with ASD and typically-developing children in the word learning context, and that this intentional reading of gaze may benefit the recollection of an object's semantic features. Moreover, this intention understanding may be present in typically-developing children in the action prediction context, whereas there may be more individual differences in children with ASD.

Importantly, positive implications of intention understanding were seen in the word learning context. In this context, learning with the referential gaze cue relative to the arrow cue resulted in locating the target object on average 150ms faster during test, and recalling on average 1 more semantic feature about the target object (across groups and across visits). These cue condition differences, particularly on in-depth learning, were somewhat surprising given that cue conditions were closely matched and attention to the target over the course of the video did not significantly differ between cue conditions. In addition, semantic features were collected and coded by a person blind to the cue condition and study hypotheses. Thus, these cue condition differences exhibit compelling consequences of learning new words because referential gaze was treated as an intentional cue. Notably, 4 children with ASD were excluded as outliers on the latency measure, and 8 children with ASD and 1 TD child were excluded on the semantic features measure, thus these positive benefits of learning with referential gaze did not include all children in the sample, particularly all children with ASD. Though mixed model analyses took into account the random effect of participants, promoting generalizability of these results beyond this sample, this sample is still relatively small and is only the first to demonstrate this benefit of with gaze relative to an arrow cue. Future studies will determine the reliability of these findings in other samples of children with ASD or typical development. These cue condition differences highlight that in the word learning context, the speed of processing and in-depth learning of recalling semantic features may be stronger because of an intentional reading of gaze in this sample of children with ASD and typically-developing children.

In the word learning context, performance on the in-depth measures suggest that children were in fact learning a word, which many studies had proposed even though often the only aspect of learning measured was selecting the object associated with the label. For example, learning on measures of semantic features and word generalization indicate that children are implicitly learning features about an object and can generalize the label to other exemplars of the target object. Though other measures of semantic knowledge such as word association were more difficult for children, there was a wide variation of individual differences (depicted in Appendix H for word associations and Appendix I for word descriptions), where many were able to provide some semantic responses for target labels. Moreover, the word description measure of semantic features was sensitive enough to detect a benefit of learning from referential gaze relative to an arrow cue. Evidence of children encoding more than a label-object association was also seen when tracking the consistency of learning across expressive language measures. A similar number of children with ASD and typically-developing children were consistently demonstrating, for a minimum of one target word, in-depth knowledge across expressive language measures, although one week later more children were demonstrating only inconsistent or no learning across measures. Taken together, these findings demonstrate that when children with ASD or typical development learn from fast-mapping with either cue, they are not only encoding a label-object association, but are beginning to encode a word.

Furthermore, on some in-depth learning measures, the evaluation of learning after a one-week delay demonstrated long-term memory consolidation (McGregor, Licandro, Arenas, Eden, Stiles, Bean, et al., 2013a). The types of measures that resulted in consolidation over a week were different than those that did not show this consolidation. That is, word association and word description measures showed weaker consolidation effects over time, and these were in-depth learning measures where children were asked to recall semantic detail about the object. Thus, the long term retention of semantic meaning, as defined in this study, may require more than teaching single word labels in scenarios with minimal information (i.e., only a referential gaze or an arrow cue). However, the other two in-depth learning measures of word generalization and word production assessed children's understanding of how to use the label itself rather than recalling semantic detail. Word generalization and word production both went beyond the initial receptive understanding of a label associated with a specific object, and examined if children could be flexible in how they generalize the label to exemplars of the target object and verbally produce the label when they see the target object.

On word generalization, children's accuracy was high immediately after the video, and remained high in both groups one week later. Moreover, children were significantly faster at responding to the original target object image and its exemplars at visit 2 versus visit 1, indicating stronger memory of the images though it is unclear if this is for target objects specifically. Additionally, though word production was difficult for both groups, the median number of correct target labels slightly increased from around 1 to 2 labels by visit 2, and more children in both groups remembered at least one correct label by visit 2. This consolidation supports other studies of improved consolidation over time, which some have attributed to sleep (Dumay & Gaskell, 2007; Henderson et al., 2012). However, it is also possible that the testing session at the one-week delay stimulated a reconsolidation "test effect", where the repeated testing resulted in reconsolidation effects that are argued to be different from true consolidation (Roediger III & Karpicke, 2006; McGregor, Licandro, Arenas, Eden, Stiles, Bean, et al., 2013a). The positive consolidation findings in the current study conflict with others who have reported impairments in children with ASD for the integration of word knowledge after a period of 24 hours (Henderson et al., 2014). It may be that word knowledge is stronger after a longer period of delay than 24 hours, as seen in the current study. Overall, these findings provide a more comprehensive picture of how children are beginning to store a fast-mapped label-object association as a word both immediately after learning and after a one-week delay.

In contrast to the word learning context, any potential intention understanding appeared to be less robust in the action prediction context. Though as seen in the word learning context, typically-developing children as a group demonstrated more attention to the area of the cue in the gaze relative to the arrow condition, this cue condition difference was not significant in children with ASD. However, the similar direction of the effect in both groups of children point to potentially heterogeneous intention understanding abilities in children with ASD rather than a deficit in this context. Yet no benefits of a potentially intentional reading of gaze were seen in either group of children on the pointing measure or on a measure that tested whether they generalized to a different color 2D block set. With both cues, it seemed that children had difficulty with this task, and were unable to overcome their knowledge of a conventional shape to use cues indicating the target block (to create an unconventional shape), despite preferences for the target relative to the distractor block during baseline and teaching. However, on the practice set of blocks (using a color sequence), there was one group difference across cue conditions, where one week later some typically-developing children, relative to children with ASD, were better able to generalize learning by choosing the 2D target block that was a different color than the tower (2 ASD, 8 TD). Though the majority of children in both groups still choose the distractor block (18 ASD, 15 TD), this group difference lends some support to Vivanti et al. (2011), who found stronger action prediction in typically-developing children relative to children with ASD. Future studies in the action prediction context must continue to examine the role of the directional or potentially intentional reading of referential gaze when children can demonstrate successful action prediction.

Word learning and action prediction are both contexts where children need to attend to a cue to help them learn. Yet many children did not have similar success with both cues in both contexts, as children with ASD or typical development had more success with both cues in the word learning context and more children in both groups were at chance or had limited or no success in the action prediction context. However, it is important to note that albeit a small number, a similar proportion of children in both groups were successful with both cues across both contexts (18% ASD, 26% TD), indicating that success with both cues in both contexts was seen in children with ASD who were matched to typically-developing children on nonverbal IQ, chronological age, ratio of girls to boys, and parental education. Although in Chapter 2 I discussed many reasons for why the action prediction context was more difficult, another reason for this difficulty could be because a word learning context is more familiar to children. The lack of familiarity with the action prediction context may make it more difficult for children to discern what is expected when a cue indicates a block. Future studies should continue to find ways to examine the role of referential gaze across different contexts, because this would provide a more complete understanding of how intention understanding is used across multiple contexts in our everyday lives.

The main contribution of this dissertation was to directly test whether children with ASD and typically-developing children use referential gaze to contribute to learning because children treat it is a directional or intentional cue. Furthermore, I evaluated in-depth learning immediately after the video as well as retention of learning one week later, with the retention of learning being a little-studied but critical element to employ new knowledge in daily life. The evaluation of cue condition differences on measures of visual attention and in-depth learning was dependent on stringent matching of the cues, to isolate whether differences between cue conditions could be specific to an intentional reading of gaze. This stringent matching may have made the videos both easy or difficult for children. The videos may have been easy in that there was only one cue to follow with limited other information to attend to, yet this also made this task much more difficult than prior studies in both contexts, which included other body cues such as head turn, voice modulation, and body positioning to highlight the target object (Bani Hani et al., 2012; Baron-Cohen et al., 1997; Luyster & Lord, 2009; Parish-Morris et al., 2007; Vivanti et al. 2011). However, these stringent methods are necessary, given that intention understanding is not directly observable (Huang et al., 2002; Premack & Woodruff, 1978), thus making it challenging to make definitive claims. Particularly in the field of ASD, it is critical to test assumptions of intention understanding in the same way in both children with ASD and control groups, rather than relying on between-group differences that cannot speak to whether there is potential variation in an intentional reading of gaze.

Gaze perception involves a large network, with some defining areas including the amygdala, the superior temporal gyrus (STG), the superior temporal sulcus (STS), and other ventro-temporal and frontal areas (e.g., for reviews on neural processing of gaze perception see Birmingham & Kingstone, 2009; Itier & Batty, 2009). In contrast to the findings of this study, studies examining neural activations have proposed that individuals with ASD do have difficulty with referential gaze processing relative to an arrow cue, with atypical neural activations in individuals with ASD (Hanson, Hanson, Ramsey, & Glymour, 2013; Vaidya et al., 2011). These findings suggest that the difficulty of gaze processing is related to the social impairments seen in individuals with ASD. However, researchers in typical development include those who either support the view of unique cortical activations for referential gaze and arrow cues (e.g., Hietanen et al., 2006; Marotta, Lupiáñez, Martella, & Casagrande, 2012b; Ristic et al., 2002), or others who suggest cortical activations are not different between these cues (e.g., Brignani, Guzzon, Marzi, & Miniussi, 2009; Tipper, Handy, Giesbrecht, & Kingstone, 2008). Thus there is yet to be a consensus on whether gaze is processed differently from an arrow in both children with ASD or typical development, and future studies must work across disciplines to better understand the connection between the brain and how children learn from this cue.

Whether we process information related to people, such as referential gaze, using lean attentional and spatial processes or rich processes of understanding another's mental state, is an ongoing debate. Different researchers have proposed their variations of these contrasting ideas using different terminology. For example, Heyes (2014) introduced the terms submentalizing versus mentalizing, where *submentalizing* refers to processes that are actually due to more general mechanisms such as attention but can falsely appear to be because of *mentalizing*, or thinking about another's mental states. Leekam uses the term *'non-social' cognition* to refer to domain-general processes that are due to perceptual and sensory cognition, whereas *'social' cognition* refers to processes that are specific to the social domain. Moreover, with respect to processing another's direction of gaze, researchers use different terms that also contrast lean and rich interpretations: Driver (1999) used the terms mechanistic versus mentalistic, where *mechanistic* refers to simply following gaze whereas *mentalistic* refers to treating gaze as intentional, and Ristic (2005) used the terms *feature correspondence* for when individuals are attending to basic perceptual features and how they correspond to each other versus a *social reading hypothesis* when individuals are reading gaze as a cue that conveys social information such as another's interest and joint attention. While many will likely continue to contribute to this debate, it is important that we continue to be rigorous in how we compare whether findings may truly be related to thinking about another's mental state or lean processes that do not require this mental understanding.

In summary, it is important to understand how both children with ASD and those with typical development use intention reading, especially since this intention reading may hold implications for learning about the world around them. As we continue to rule out alternative interpretations (Heyes, 2014b; 2016), we may be able to pinpoint the role of intention understanding in individuals with ASD and those with typical development.

Limitations and Future Directions

One limitation to this dissertation was the number of trials per cue condition. The number of trials was chosen because it was consistent with prior studies on learning in these contexts (Bani Hani et al., 2012; Luyster & Lord, 2009; Parish-Morris et al., 2007; Vivanti et al., 2011), but also because both contexts were included. Because including both contexts meant that children were watching double the number of videos than had been seen in prior studies, I did not want to add additional trials to each context. Adding more trials may have affected the quality of learning in both contexts, but particularly in the word learning context when I wanted to assess children's in-depth learning. Future studies that focus on one context should include more trials to better assess cue condition differences.

Another limitation of this study was the lack of including the full range of children with ASD on the spectrum. I included only children within the normal or above average range of IQ, and did not include those with more severe intellectual delays, which excluded approximately half of children on the spectrum (Charman et al., 2011). Therefore, it is unclear whether these results can be generalized to other children with ASD who do not share the characteristics of those in my sample. This heterogeneity within children with ASD is a major challenge in this population, because it limits what can be interpreted to the population as a whole. One way to address this issue is by acknowledging the individual differences that can contribute to a particular skill by examining the role of covariates, rather than focusing on the outcome itself (Kover & Atwood, 2013). However, because of the small sample size, neither chapter was able to address both cue condition effects and the covariates that may be associated with individual variability in children's performances in both contexts. For example, language abilities and IQ have been shown to be predictors of children's visual attention to referential gaze in other studies (Falck-Ytter et al., 2012; Leekam et al., 1998).

Furthermore, another factor that was not examined in either group was the role of children's visual memory, which may have affected children's learning immediately after the video and over time. For example, in the word learning context, children with stronger visual memories in both groups may have been the ones who remembered more semantic features, or, in the action prediction context, children with stronger visual memories may have been the ones to successfully predict the target block. Yet having a stronger visual memory would not have diminished the finding of more semantic features provided in the gaze over the arrow condition. Improved visual memory skills could have increased semantic features similarly with both cues. Nevertheless, it is critical to understand the factors that contribute to intention understanding from referential gaze.

Other future directions for this study include adding control groups (e.g., children with other neurodevelopmental disorders such as specific language impairment or ADHD) and adding more cues. Additional control groups are important to address questions of how intention understanding is used across populations. For example, if intention understanding in the context of action prediction is heterogeneous in children with ASD when it was not seen to be as heterogeneous in typically-developing children, including a control group in this context could examine whether this heterogeneity is specific to children with ASD and thus related to their social impairments or whether this heterogeneity may be seen in other clinical populations and thus may not be due to a social impairment per se.

Moreover, it is important to go beyond a single referential gaze shift to understand how other cues work with referential gaze, as in our more natural everyday interactions. For example, emotion cues may play an important role to confirm situations of ambiguity such as in the case of action prediction in this study. Because of the strong conventional pull of the distractor block, perhaps children needed both a referential gaze shift and a smile to the unconventional target block to better solidify the person’s intention to use this block to build the tower.

General Conclusion

One of the ways that children learn about the world around them is through their interactions with people. Though people differ in exactly how they interact, many of the same set of communication tools are at their disposable, such as their referential gaze. Children with ASD have known difficulties in their social communication and interaction abilities, which may put them at a significant disadvantage when they are expected to make use of such subtle cues to help them learn. However, in this dissertation I found that both children with ASD without intellectual delay and typically-developing children similarly used an intentional reading of gaze to learn new words, and this intentional reading benefited their in-depth learning of semantic features immediately after the video and one week later. Future research will need to better elucidate the role of intention understanding in the context of action prediction. These findings are the first to provide direct evidence to support an intentional, not a directional, reading of referential gaze a word learning context, which was similarly found in both children with ASD and typically-developing children. For all future researchers, I hope to have highlighted an area of study that is limitless in feeding one's curiosity, but can also challenge one's creativity and patience.

REFERENCES

Akechi, H., Kikuchi, Y., Tojo, Y., & Osanai, H. (2013). Brief report: Pointing cues facilitate word learning in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *43*(1), 230–235. http://doi.org/10.1007/s10803-012-1555-3

Akechi, H., Senju, A., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2011). Do children with ASD use referential gaze to learn the name of an object? An eye-tracking study. *Research in Autism Spectrum Disorders*, *5*(3), 1230–1242. http://doi.org/10.1016/j.rasd.2011.01.013

Aldaqre, I., Paulus, M., & Sodian, B. (2015). Referential gaze and word learning in adults with autism. *Autism Research*, *19*(8), 944–955. http://doi.org/10.1177/1362361314556784

Aldridge, M. A., Stone, K. R., Sweeney, M. H., & Bower, T. (2000). Preverbal children with autism understand the intentions of others. *Developmental Science*, *3*(3), 294–301.

Allen, C. W., Silove, N., Williams, K., & Hutchins, P. (2007). Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders*, *37*(7), 1272–1278. http://doi.org/10.1007/s10803-006-0279-7

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, *137*(2), 190–200. http://doi.org/10.1016/j.actpsy.2010.09.009

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Ames, C. S., & Jarrold, C. (2006). The problem with using eye-gaze to infer desire: a deficit of cue inference in children with autism spectrum disorder? *Journal of Autism and*

Developmental Disorders, 37(9), 1761–1775.

Archibald, L. M. D. L., & Joanisse, M. F. M. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, *52*(4), 899–914. http://doi.org/10.1044/1092-4388(2009/08-0099)

Astington, J., & Jenkins, J. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*(5), 1311-1320.

Baldwin, D. A. (1993a). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, *29*(5), 832–843.

Baldwin, D. A. (1993b). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(2), 395–418.

Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, *5*(4), 171–178.

Baldwin, D., Markman, E., Bill, B., Desjardins, R., Irwin, J., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development*, *67*(6), 3135–3153.

Bani Hani, H., Gonzalez-Barrero, A. M., & Nadig, A. S. (2012). Children's referential understanding of novel words and parent labeling behaviors: Similarities across children with and without autism spectrum disorders. *Journal of Child Language*, *40*(5), 971–1002. http://doi.org/10.1017/S0305000912000426

Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Baldwin, D., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development*, *68*(1), 48–57.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology (I995)*, *13*, 379–398.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. http://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models. *Manuscript submitted for publication*, 1–27.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. http://doi.org/10.18637/jss.v067.i01

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, *12*(7), 254–258. http://doi.org/10.1016/j.tics.2008.04.005

Becchio, C., Pierno, A., M, M., Lusher, D., & Castiello, U. (2007). Motor contagion from gaze: the case of autism. *Brain*, *130*(9), 2401–2411. http://doi.org/10.1093/brain/awm171

Bellagamba, F., & Tomasello, M. (1999). Re-enacting intended acts: Comparing 12-and 18-month-olds. *Infant Behavior and Development*, *22*(2), 277–282.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York, NY: John Wiley & Sons, Inc. http://doi.org/10.1002/0471725153.ch2

Benjamin, D. P., McDuffie, A. S., Thurman, A. J., Kover, S. T., Mastergeorge, A. M., Hagerman, R. J., & Abbeduto, L. (2015). Effect of speaker gaze on word learning in fragile X syndrome: A comparison with nonsyndromic autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, *58*(2), 383-395. http://doi.org/10.1044/2015_JSLHR-L-14-0136

Birmingham, E., & Kingstone, A. (2009). Human Social Attention. *Annals of the New York Academy of Sciences*, *1156*(1), 118–140. http://doi.org/10.1111/j.1749-6632.2009.04468.x

Birmingham, E., Ristic, J., & Kingstone, A. (2012). *Investigating Social Attention*. J. A. Burack, J. T. Enns, & N. A. Fox (Eds.). New York, NY: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780195315455.001.0001

Blackford, J. U. (2007). Statistical issues in developmental epidemiology and developmental disabilities research: Confounding variables, small sample size, and numerous outcome variables. *International Review of Research in Mental Retardation*, *33*, 93–120. http://doi.org/10.1016/S0074-7750(06)33005-4

Blackford, J. U. (2009). Propensity scores: Method for matching on multiple variables in down syndrome research. *Intellectual and Developmental Disabilities*, *47*(5), 348–357. http://doi.org/10.1352/1934-9556-47.5.348

Boraston, Z., & Blakemore, S. J. (2007). The application of eye-tracking technology in the study of autism. *The Journal of Physiology*, *581*(3), 893–898. http://doi.org/10.1113/jphysiol.2007.133587

Brignani, D., Guzzon, D., Marzi, C. A., & Miniussi, C. (2009). Attentional orienting induced by arrows and eye-gaze compared with an endogenous cue. *Neuropsychologia*, *47*(2), 370–381. http://doi.org/10.1016/j.neuropsychologia.2008.09.011

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*, 55–72.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language and Development*, *15*, 17–29.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*(2), 315–330.

Carpenter, M., Pennington, B., & Rogers, S. (2001). Understanding of others' intentions in children with autism. *Journal of Autism and Developmental Disorders*, *31*(6), 589–599.

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological Medicine*, *41*(03), 619–627. http://doi.org/10.1017/S0033291710000991

Chawarska, K., Macari, S., & Shic, F. (2012). Context modulates attention to social scenes in toddlers with autism. *Journal of Child Psychology and Psychiatry*, *53*(8), 903–913. http://doi.org/10.1111/j.1469-7610.2012.02538.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Colombi, C., Liebal, K., Tomasello, M., Young, G., Warneken, F., & Rogers, S. J. (2009). Examining correlates of cooperation in autism: Imitation, joint attention, and understanding intentions. *Autism Research*, *13*(2), 143–163. http://doi.org/10.1177/1362361308098514

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(6), 741–748.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Jr., Leventhal, B. L., & Lord, C. (2007). Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry*, *48*(9), 932–940. http://doi.org/10.1111/j.1469-7610.2007.01762.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153.

D'Entremont, B., & Yazbek, A. (2006). Imitation of intentional and accidental actions by children with autism. *Journal of Autism and Developmental Disorders*, *37*(9), 1665–1678. http://doi.org/10.1007/s10803-006-0291-y

Del Re, A.C. (2013). compute.es: Compute effect sizes. R package version 0.2-2. Retrieved from http://cran.r-project.org/web/packages/compute.es

Doherty, M. J. (2006). The development of mentalistic gaze understanding. *Infant and Child Development*, *15*(2), 179–186. http://doi.org/10.1002/icd.434

Doherty, M. J., McIntyre, A. H., & Langton, S. R. H. (2015). Developmentally distinct gaze processing systems: Luminance versus geometric cues. *Cognition*, *137*, 72–80. http://doi.org/10.1016/j.cognition.2015.01.001

Driver, J., IV, Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*(5), 509–540. http://doi.org/10.1080/135062899394920

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35–39. http://doi.org/10.1111/j.1467-9280.2007.01845.x

Duyck, W., Desmet, T., Verbeke, L. P., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods*, *36*(3), 488–499.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*, 581–604.

Falck-Ytter, T., Fernell, E., Hedvall, Å. L., Hofsten, von, C., & Gillberg, C. (2012). Gaze performance in children with autism spectrum disorder when observing communicative actions. *Journal of Autism and Developmental Disorders*, *42*(10), 2236–2245. http://doi.org/10.1007/s10803-012-1471-6

Farroni, T., Johnson, M. H., Brockbank, M., & Simion, F. (2000). Infants' use of gaze direction to cue attention: The importance of perceived motion. *Visual Cognition*, *7*(6), 705–718. http://doi.org/10.1080/13506280050144399

Fenson, L., Marchman, V., Thal, D., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore: Paul H Brookes Publishing.

Fernald, A., Perfors, A., & Marchman, V. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116.

Field, C. (2016). Word learning from object, speaker, and environmental cues in typically developing children, children with autism spectrum disorder and children with other developmental disorders (Doctoral dissertation). Retrieved from http://eprintslancsacuk/78199/.

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. http://doi.org/10.1016/j.neuropsychologia.2008.07.016

Gamer, M., Lemon, J., Fellows, I., & Puspendra, S. (2012). irr: Various coefficients of interrater reliability and agreement. R package version 0.84. Retrieved from https://CRAN.R-project.org/package=irr

Gillespie-Lynch, K., Elias, R., Escudero, P., Hutman, T., & Johnson, S. P. (2013). Atypical gaze following in autism: A comparison of three potential mechanisms. *Journal of Autism and Developmental Disorders*, *43*(12), 2779–2792. http://doi.org/10.1007/s10803-013-1818-7

Gladfelter, A. (2014). The role of semantic richness in word learning in children with autism spectrum disorder, specific language impairment, and typical language development (Doctoral dissertation). Retrieved from http://gradworks.umi.com/36/69/3669227.html, 1–141.

Gliga, T., Elsabbagh, M., Hudry, K., Charman, T., Johnson, M. H., The BASIS Team. (2012). Gaze following, gaze reading, and word learning in children at risk for autism. *Child Development*, *83*(3), 926–938. http://doi.org/10.1111/j.1467-8624.2012.01750.x

Gredebäck, G., Johnson, S., & Hofsten, von, C. (2009). Eye tracking in infancy research. *Developmental Neuropsychology*, *35*(1), 1–19. http://doi.org/10.1080/87565640903325758

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34.

Hanson, C., Hanson, S. J., Ramsey, J., & Glymour, C. (2013). Atypical effective connectivity of social brain networks in individuals with autism. *Brain Connectivity*, *3*(6), 578–589.

http://doi.org/10.1089/brain.2013.0161

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, *66*(3), 843–855.

Henderson, L. M., Weighall, A. R., Brown, H., & Gareth Gaskell, M. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, *15*(5), 674–687. http://doi.org/10.1111/j.1467-7687.2012.01172.x

Henderson, L., Powell, A., Gareth Gaskell, M., & Norbury, C. (2014). Learning and consolidation of new spoken words in autism spectrum disorder. *Developmental Science*, *17*(6), 858–871. http://doi.org/10.1111/desc.12169

Heyes, C. (2014a). False belief in infancy: a fresh look. *Developmental Science*, *17*(5), 647–659. http://doi.org/10.1111/desc.12148

Heyes, C. (2014b). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*(2), 131–143. http://doi.org/10.1177/1745691613518076

Heyes, C. (2016). Who knows? Metacognitive social learning strategies. *Trends in Cognitive Sciences*, *20*(3), 204–213. http://doi.org/10.1016/j.tics.2015.12.007

Hietanen, J. K., Nummenmaa, L., Nyman, M. J., Parkkola, R., & Hämäläinen, H. (2006). Automatic attention orienting by social and symbolic cues activates different neural networks: An fMRI study. *NeuroImage*, *33*(1), 406–413. http://doi.org/10.1016/j.neuroimage.2006.06.048

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. http://doi.org/10.1093/pan/mpl013

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1-28.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88.

Hollich, G., Hirsh-Pasek, K., Golinkoff, R., Brand, R., Brown, E., Chung, H., et al. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, *65*(3), i-135.

Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science*, *12*(5), 360–365.

Howarth, K. (2010). *Word learning and vocabulary growth in children with autism spectrum disorder* (Unpublished masters thesis). McGill University, Montreal, Canada.

Huang, C-T., & Charman, T. (2005). Gradations of emulation learning in infants' imitation of actions on objects. *Journal of Experimental Child Psychology*, *92*(3), 276–302. http://doi.org/10.1016/j.jecp.2005.06.003

Huang, C-T., Heyes, C., & Charman, T. (2002). Infants' behavioral reenactment of 'failed attempts': Exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental Psychology*, *38*(5), 840–855. http://doi.org/10.1037//0012-1649.38.5.840

Huang, C-T., Heyes, C., & Charman, T. (2006). Preschoolers' behavioural reenactment of "failed attempts": The roles of intention-reading, emulation and mimicry. *Cognitive Development*, *21*(1), 36–45. http://doi.org/10.1016/j.cogdev.2005.09.002

Hurtado, N., Marchman, V., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), 31–39. http://doi.org/10.1111/j.1467-7687.2008.00768.x

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews*, *33*(6), 843–863. http://doi.org/10.1016/j.neubiorev.2009.02.004

King, G., & Nielsen, R. (2016). *Why propensity scores should not be used for matching.* Manuscript in preparation. Retrieved from http://j.mp/1sexgVw

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes*, *16*(2-3), 287–308. http://doi.org/10.1080/01690960042000058

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*(6635), 767–768. http://doi.org/10.1038/42842

Kover, S. T., & Atwood, A. K. (2013). Establishing equivalence: Methodological progress in group-matching design and analysis. *American Journal on Intellectual and Developmental Disabilities*, *118*(1), 3–15. http://doi.org/10.1352/1944-7558-118.1.3

Kreger, K. L. (2016). *The production of semantic representations in children with specific language impairment, autism spectrum disorder, and typical language development* (Masters thesis). Retrieved from http://gradworks.umi.com/10/12/10123496.html, 1–122.

Kuhn, G., Benson, V., Fletcher-Watson, S., Kovshoff, H., McCormick, C. A., Kirkby, J., & Leekam, S. R. (2010). Eye movements affirm: automatic overt gaze and arrow cueing for typical adults and adults with autism spectrum disorder. *Experimental Brain Research*, *201*(2), 155–165. http://doi.org/10.1007/s00221-009-2019-7

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-32. Retrieved from https://CRAN.R-project.org/package=lmerTest

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(863), 1–12. http://doi.org/10.3389/fpsyg.2013.00863

Leekam, S. (2016). Social cognitive impairment and autism: What are we trying to explain? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150082.

Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and cues: Gaze-following in children with autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *39*(7), 951–962.

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, *15*(1), 77–95.

Leighton, J., Bird, G., & Heyes, C. (2010). “Goals” are not an integral component of imitation. *Cognition*, *114*(3), 423–435. http://doi.org/10.1016/j.cognition.2009.11.001

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1), 1-33. http://doi.org/10.18637/jss.v069.i01

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

Loveland, K. A., & Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *Journal of Autism and Developmental Disorders*, *16*(3), 335–349.

Luyster, R., & Lord, C. (2009). Word learning in children with autism spectrum disorders. *Developmental Psychology*, *45*(6), 1774–1786. http://doi.org/10.1037/a0016223

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., et al. (2009). The Autism diagnostic observation schedule—Toddler module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(9), 1305–1320. http://doi.org/10.1007/s10803-009-0746-z

Marotta, A., Lupiáñez, J., & Casagrande, M. (2012a). Investigating hemispheric lateralization of reflexive attention to gaze and arrow cues. *Brain and Cognition*, *80*(3), 361–366. http://doi.org/10.1016/j.bandc.2012.08.001

Marotta, A., Lupiáñez, J., Martella, D., & Casagrande, M. (2012b). Eye gaze versus arrows as spatial cues: Two qualitatively different modes of attentional selection. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 326–335. http://doi.org/10.1037/a0023959

Marsh, L. E., & de C Hamilton, A. F. (2011). Dissociation of mirroring and mentalising systems in autism. *NeuroImage*, *56*(3), 1511–1519. http://doi.org/10.1016/j.neuroimage.2011.02.003

Marsh, L. E., Pearson, A., Ropar, D., & Hamilton, A. F. de C. (2015). Predictive gaze during observation of irrational actions in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *45*(1), 245–261. http://doi.org/10.1007/s10803-014-2215-6

McDuffie, A., Yoder, P., & Stone, W. (2006). Fast-mapping in young children with autism spectrum disorders. *First Language*, *26*(4), 421–438. http://doi.org/10.1177/0142723706067438

McGregor, K. K., Berns, A. J., Owen, A. J., Michels, S. A., Duff, D., Bahnsen, A. J., & Lloyd, M. (2012). Associations between syntax and the lexicon among children with or without asd and language impairment. *Journal of Autism and Developmental Disorders*, *42*, 35–47.

http://doi.org/10.1007/s10803-011-1210-4

McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013a). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1845–1856. http://doi.org/10.1044/1092-4388(2013/12-0233)

McGregor, K. K., Rost, G., Arenas, R., Farris-Trimble, A., & Stiles, D. (2013b). Children with ASD can use gaze in support of word recognition and learning. *Journal of Child Psychology and Psychiatry*, *54*(7), 745–753. http://doi.org/10.1111/jcpp.12073

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838.

Meltzoff, A. N. (2007). "Like me": a foundation for social cognition. *Developmental Science*, *10*(1), 126–134. http://doi.org/10.1111/j.1467-7687.2007.00574.x

Mervis, C. B., & Robinson, B. F. (1999). Methodological issues in cross-syndrome comparisons: Matching procedures, sensitivity (Se), and specificity (Sp). *Monographs of the Society for Research in Child Development*, *64*(1), 115–130.

Mervis, C., & Klein-Tasman, B. (2004). Methodological issues in group-matching designs: α levels for control variable comparisons and measurement characteristics of control and target variables. *Journal of Autism and Developmental Disorders*, *34*(1), 7–17.

Moore, C., & Corkum, V. (1994). Social understanding at the end of the first year of life. *Developmental Review*, *14*, 349–372.

Moore, C., & Corkum, V. (1998). Infant gaze following based on eye direction. *British Journal of Developmental Psychology*, *16*(4), 495–503.

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology*, *35*(1), 60–68.

Nadig, A., & Bang, J. (2016). Parental input to children with ASD and its influence on later language. In L. Naigles (Ed.), *Innovative Investigations of Language in Autism Spectrum Disorder* (pp. 89-114). Washington, DC: APA Books.

Nadig, A., & Mulligan, A. (in press). Intact non-word repetition and similar error patterns in language-matched children with autism spectrum disorders: A pilot study. *Journal of Communication Disorders*.

Nadig, A., Vivanti, G., & Ozonoff, S. (2009). Adaptation of object descriptions to a partner under increasing communicative demands: A comparison of children with and without autism. *Autism Research*, *2*, 334–347. http://doi.org/10.1002/aur.102

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? Is it automatic? Is it social? *Development and Psychopathology*, *20*(01), 79–97. http://doi.org/10.1017/S0954579408000047

Nieuwenhuis, R., te Grotenhuis, H. F., & Pelzer, B. J. (2012). Influence.ME: Tools for detecting influential data in mixed effects models. *The R Journal*, *4*(2), 38–47.

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 and Psychiatry*, *50*(7), 834–842. http://doi.org/10.1111/j.1469-7610.2009.02073.x

Norbury, C. F., Griffiths, H., & Nation, K. (2010). Sound before meaning: Word learning in autistic disorders. *Neuropsychologia*, *48*, 4012–4019.

Oakes, L. M. (2012). Advances in eye tracking in infancy research. *Infancy*, *17*(1), 1–8. http://doi.org/10.1111/j.1532-7078.2011.00101.x

Parish-Morris, J., Hennon, E. A., Hirsh-Pasek, K., Golinkoff, R. M., & Tager-Flusberg, H. (2007). Children with autism illuminate the role of social intention in word learning. *Child Development*, *78*(4), 1265–1287.

Pickles, A., Anderson, D. K., & Lord, C. (2014). Heterogeneity and plasticity in the development of language: A 17-year follow-up of children referred early for possible autism. *Journal of Child Psychology and Psychiatry*, *55*(12), 1354–1362. http://doi.org/10.1111/jcpp.12269

Pierno, A., M, M., Glover, S., Georgiou, I., & Castiello, U. (2006). Failure to read motor intentions from gaze in children with autism. *Neuropsychologia*, *44*(8), 1483–1488. http://doi.org/10.1016/j.neuropsychologia.2005.11.013

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition*, *97*(1), B13–B23. http://doi.org/10.1016/j.cognition.2005.01.008

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *4*, 515–526.

Pruett, J. R., LaMacchia, A., Hoertel, S., Squire, E., McVey, K., Todd, R. D., et al. (2010). Social and non-social cueing of visuospatial attention in autism and typical development. *Journal of Autism and Developmental Disorders*, *41*(6), 715–731. http://doi.org/10.1007/s10803-010-1090-z

Quillian, M. R. (1969). The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM*, *12*(8), 459–476.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Revelle, W. (2016, November 23). psych: Procedures for personality and psychological research, Northwestern University, Evanston, Illinois, USA. Retrieved from http://CRAN.R-project.org/package=psych

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review*, *9*(3), 507–513.

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research*, *24*(3), 715–718. http://doi.org/10.1016/j.cogbrainres.2005.02.007

Ristic, J., Wright, A., & Kingstone, A. (2007). Attentional control and reflexive orienting to gaze and arrow cues. *Psychonomic Bulletin & Review*, *14*(5), 964–969.

Roediger III, H. L. R., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Roid, G. H., & Miller, L. J. (2013). *Leiter International Performance Scale – Third Edition*. Wood Dale, IL: Stoelting.

Rombough, A., Barrie, J. N., & Iarocci, G. (2012). Social cueing elicits a distinct form of visual spatial orienting. In J. A. Burack, J. T. Enns, & N. A. Fox (Eds.), *Cognitive Neuroscience, Development, and Psychopathology: Typical and Atypical Developmental Trajectories of Attention* (221–250). New York, NY: Oxford University Press.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169–188.

Rutter, M., Bailey, A., & Lord, C. (2003). *Social communication questionnaire (SCQ)*. Los Angeles: Western Psychological Services.

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266.

Secord, W., Wiig, E., Boulianne, L., Semel, E., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales - Version pour francophones du Canada*. Toronto: Pearson Canada Assessment.

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals - 4th Edition*. *The Psychological Corporation*. San Antonio, TX: The Psychological Corporation.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in asperger syndrome. *Science*, *325*(5942), 883–885. http://doi.org/10.1126/science.1176170

Senju, A., Tojo, Y., Dairoku, H., & Hasegawa, T. (2004). Reflexive orienting in response to eye gaze and an arrow in children with and without autism. *Journal of Child Psychology and Psychiatry*, *45*(3), 445–458.

Sheng, L., & McGregor, K. K. (2010). Lexical-semantic organization in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *53*(1), 146–159.

Sparrow, S. S., Cicchetti, D., & Balla, D. A. (2005). *Vineland adaptive behavior scales - 2nd edition*. Minneapolis, MN: NCS Pearson, Inc.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. http://doi.org/10.1214/09-STS313

Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs: Matching methods for causal inference. In J. Osborne (Ed.), *Best Practices in Quantitative Social Science* (pp. 155–176). Thousand Oaks, CA: Sage Publications.

Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clinical Neuroscience*

Research, *6*(3-4), 219–224.

Tager-Flusberg, H., & Joseph, R. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1430), 303.

Tenenbaum, E. J., Amso, D., Abar, B., & Sheinkopf, S. J. (2014). Attention and word learning in autistic, language delayed and typically developing children. *Frontiers in Psychology*, *5*, 490, 1-9. http://doi.org/10.3389/fpsyg.2014.00490

Thurm, A., Lord, C., Lee, L., & Newschaffer, C. (2007). Predictors of language acquisition in preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(9), 1721–1734.

Tipper, C. M., Handy, T. C., Giesbrecht, B., & Kingstone, A. (2008). Brain responses to biological relevance. *Journal of Cognitive Neuroscience*, *20*(5), 879–891.

Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin & Review*, *9*(2), 314–318.

Tipples, J. (2008). Orienting to counterpredictive gaze and arrow cues. *Perception & Psychophysics*, *70*(1), 77–87. http://doi.org/10.3758/PP.70.1.77

Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, *28*, 509–529.

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, *23*(1), 157–176. http://doi.org/10.1017/S0305000900010138

Uithol, S., & Paulus, M. (2013). What do infants understand of others' action? A theoretical account of early social cognition. *Psychological Research*, *78*(5), 609–622.

http://doi.org/10.1007/s00426-013-0519-3

Vaidya, C. J., Foss-Feig, J., Shook, D., Kaplan, L., Kenworthy, L., & Gaillard, W. D. (2011). Controlling attention to gaze and arrows in childhood: An fMRI study of typical development and autism spectrum disorders. *Developmental Science*, *14*(4), 911–924. http://doi.org/10.1111/j.1467-7687.2011.01041.x

Venker, C. E., & Kover, S. T. (2015). An open conversation on using eye-gaze methods in studies of neurodevelopmental disorders. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1719–1732. http://doi.org/10.1044/2015_JSLHR-L-14-0304

Venker, C. E., Kover, S. T., & Weismer, S. E. (2016). Brief report: Fast mapping predicts differences in concurrent and later language abilities among children with ASD. *Journal of Autism and Developmental Disorders*, *46*(3), 1118–1123. http://doi.org/10.1007/s10803-015-2644-x

Vivanti, G., & Hamilton, A. (2014). Imitation in autism spectrum disorders. In F. Volkmar, R. Paul, S. J. Rogers, & K. A. Pelphrey (Eds.), *The Handbook of autism and pervasive developmental disorders 4th ed.* (pp. 278-301), New York: Wiley.

Vivanti, G., & Rogers, S. J. (2011). Action understanding and social learning in Autism: a developmental perspective. *Life Span and Disability*, *1*, 7–29.

Vivanti, G., McCormick, C., Young, G. S., Abucayan, F., Hatt, N., Nadig, A., et al. (2011). Intact and impaired mechanisms of action understanding in autism. *Developmental Psychology*, *47*(3), 841–856. http://doi.org/10.1037/a0023105

Vivanti, G., Trembath, D., & Dissanayake, C. (2014). Atypical monitoring and responsiveness to goal-directed gaze in autism spectrum disorder. *Experimental Brain Research*, *232*(2), 695–701. http://doi.org/10.1007/s00221-013-3777-9

Vlamings, P. H. J. M., Stauder, J. E. A., Son, I. A. M., & Mottron, A. L. (2005). Atypical visual orienting to gaze- and arrow-cues in adults with high functioning autism. *Journal of Autism and Developmental Disorders*, *35*(3), 267–277. http://doi.org/10.1007/s10803-005-3289-y

Walton, K. M., & Ingersoll, B. R. (2013). Expressive and receptive fast-mapping in children with autism spectrum disorders and typical development: The influence of orienting cues. *Research in Autism Spectrum Disorders*, *7*(6), 687–698. http://doi.org/10.1016/j.rasd.2013.02.012

Want, S. C., & Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of “imitation” in children. *Developmental Science*, *5*(1), 1–41.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263. http://doi.org/10.1016/j.tics.2009.03.006

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience & Biobehavioral Reviews*, *36*(3), 1060–1084. http://doi.org/10.1016/j.neubiorev.2011.12.008

Westfall, J. (2016, March 3). Five different “Cohen’s d” statistics for within-subject designs [Web log post]. Retrieved from http://jakewestfall.org/blog/index.php/2016/03/25/five-different- cohens-d-statistics-for-within-subject-designs/

Wiggins, L. D., Bakeman, R., Adamson, L. B., & Robins, D. L. (2007). The utility of the Social Communication Questionnaire in screening for autism in children referred for early intervention. *Focus on Autism and Other Developmental Disabilities*, *22*(1), 33–38.

Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*, *6*(3), 297-311.

Xu, B., & Tanaka, J. W. (2013). Does face inversion qualitatively change face processing: An eye movement study using a face change detection task. *Journal of Vision*, *13*(2), 1–16.

Appendix A. Matching Protocol

Difficulties with Quasi-Experimental Designs

The current study is a *quasi-experimental design*, meaning that participants in each group of children are not assigned randomly, but are determined because of a pre-existing characteristic that determined group assignment (e.g., presence of ASD). Because of this pre-existing characteristic, studies must match children between groups on all other known characteristics (e.g., IQ, age) to best interpret any between-group differences as being attributable to the pre-existing characteristic of interest. Issues such as selection bias can result in differences between other known characteristics that could covary with the outcome, also referred to as *covariates*. Matching aims to balance groups on these covariates and remove biases that may confound the relationship between group assignment and the outcome variable. For example, if two groups differ on IQ, then differences on an outcome variable could be attributed to either group assignment or IQ. Reducing these biases minimizes confounding effects and strengthens the independence between group assignment and the outcome variable (Blackford, 2007; Kover & Atwood, 2013).

Propensity Scores as a Matching Method

There are many different matching methods (Kover & Atwood, 2013), but I chose to match participants based on propensity scores. Propensity scores are created by summarizing multiple covariates into a single score for each participant. These scores represent an individual's conditional probability of being in a group given the specified covariates (Rosenbaum & Rubin, 1983), which are determined from prior research theory (Kover & Atwood, 2013; Stuart, 2010).

There are multiple parametric models that are used to calculate propensity scores, and the current dissertation used the method of logistic regression (Blackford, 2007; Ho et al., 2011).

There are multiple benefits with using propensity scores. For one, propensity scores avoid the subjective bias of hand selecting matches when the person making the selections may know about children's behavior during testing and characteristics. Additionally, there are often multiple covariates to match between groups, and it is difficult to match closely on all covariates (Stuart & Rubin, 2008). Therefore, because propensity scores summarize multiple covariates in a single scalar score per participant, it is presumably easier to match participants on one score than multiple. Finally, when all possible covariates are included in a propensity score, matching on propensity scores can mimic a fully randomized study (see Stuart, 2010 for further discussion on this topic). Randomization allows for a stronger interpretations of causality such that differences on the outcome variable can be attributed to true group differences. However, mimicking a fully randomized design is not the case in our small sample, because we were not able to account for all possible covariates at a time, such as language abilities. To include as many children with ASD as possible to match to TD children, I decided not to include language abilities as a covariate to match groups. Because language abilities were inherently different between groups in my sample, this difference would not be resolved by using propensity scores.

Notwithstanding these benefits, some limitations of using propensity scores include their limited use in research in neurodevelopmental disabilities (Blackford, 2007); see Blackford, 2009) for use in children with Down Syndrome), and sample size requirements. Although some authors suggest large sample sizes are required (Kover & Atwood, 2013), Blackford (2007) suggests 5 – 10 participants per covariate based on the sample size in the main group of interest (e.g., children with ASD) to be sufficient.

The following procedure details how I implemented propensity scores in the current dissertation, which was based on guidelines proposed by Stuart (2010). It is critical to match children prior to all outcome analyses, to avoid bias based on outcomes (Ho et al., 2011; Stuart, 2010). Any participants who met exclusion criteria were not included. Please see the Participants section for more details regarding exclusion criteria.

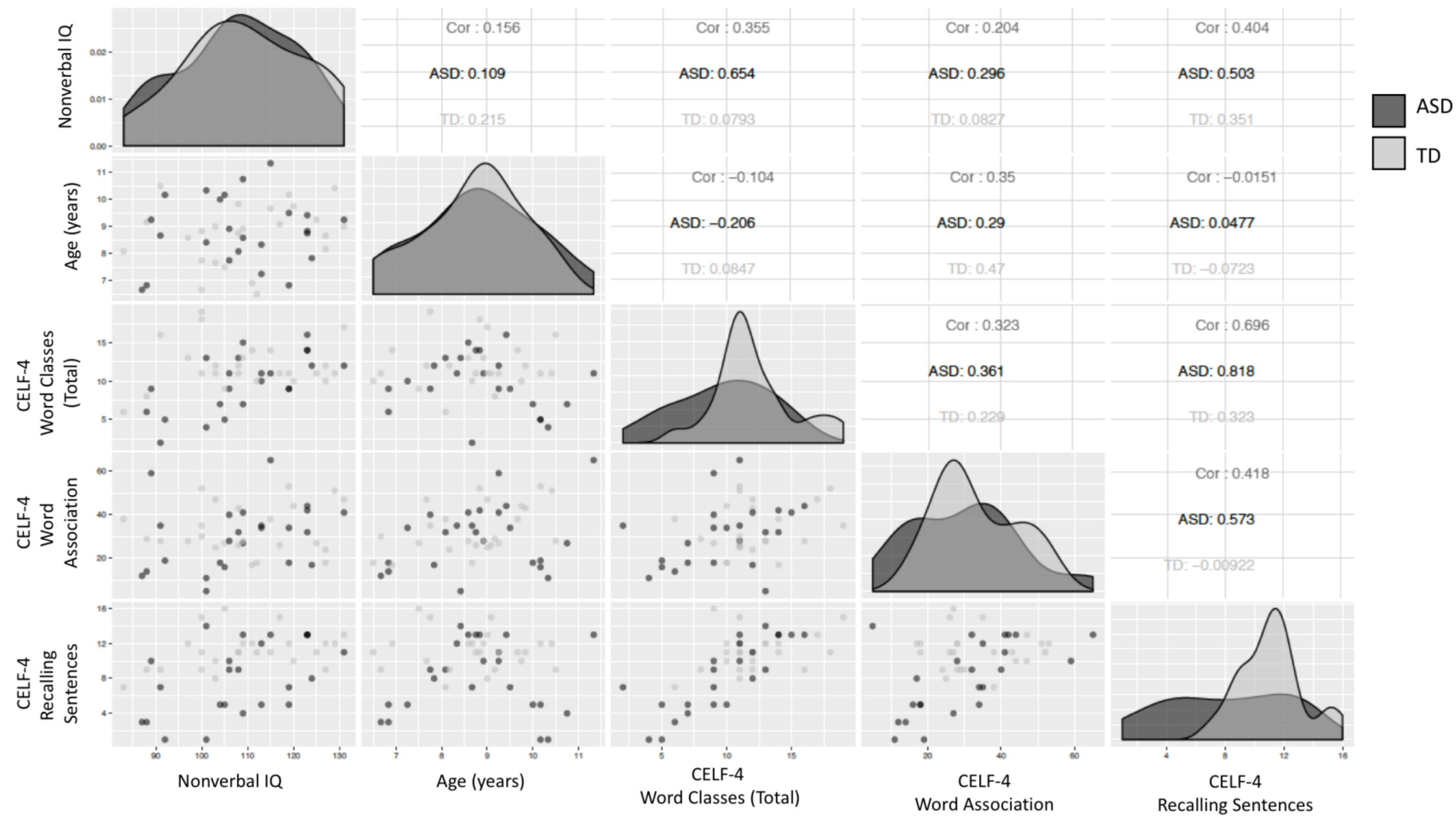
Protocol for Propensity Scores in the Current Dissertation

First, prior to data collection, I considered multiple covariates that could influence children's performance on the experimental tasks in this dissertation. These covariates were determined based on their known relationships with referential gaze following and language abilities, as well as expected relationships in the action prediction context. These covariates included: nonverbal IQ, age, ratio of girls to boys, parental education, and children's language abilities (Bani Hani et al., 2012; Baron-Cohen et al., 1997; Hoff, 2006; Leekam et al., 1998; Parish-Morris et al., 2007).

After data collection, I decided to exclude covariates of sex and parental education, because the other covariates of nonverbal IQ, age, and language abilities have been shown to be related to referential gaze following and/or language outcomes (Leekam et al., 1998; Thurm et al., 2007). I examined the distributions and interrelations between covariates of nonverbal IQ, age, and children's language abilities using the three CELF-4 subtests of language used in this dissertation, Word Classes – Total, Word Associations, and Recalling Sentences). Based on these distributions, I determined that language abilities, as measured by the three subtests (CELF 4 Word Classes, Word Associations, and Recalling Sentences) were too different between groups such that achieving adequately matched groups would result in excluding too many children with

ASD. These distributions and interrelations can be seen in Figure A 1. Whereas language abilities in TD children were within the normal range or above, in children with ASD language abilities ranged from 2 standard deviations below to 2 standard deviations above the mean. Because the primary goal of these studies was to understand the performance of children with ASD and retain natural variation in this group, I opted not to match on language abilities. Following the guidelines by Blackford (2007), I included only two covariates out of the five I had available, based on my sample size of 25 children with ASD: nonverbal IQ and age; this was roughly one covariate per 10 children. These covariates were selected because of their known relationships with referential gaze following and language outcomes in children with ASD (Leekam et al., 1998; Thurm et al., 2007). Though parental education and sex are also known for their relationships with language abilities (Fenson et al., 2007; Hoff, 2006), I prioritized nonverbal IQ and age due to their known role as important covariates in children with ASD.

Figure A 1. Matrix of Covariate Distributions and Correlations in Children with ASD and TD Children



This figure depicts the five covariates that were considered to be included in propensity scores: nonverbal IQ (measured by the Leiter), age, and language abilities measured by the CELF-4 subtests of Word Classes, Word Association, and Recalling Sentences. Through visual inspection of this matrix (and boxplots, not pictured here), I determined that the distribution of scores on language measures were inherently too different between children with ASD and TD children, thus language abilities were not included in propensity scores. The values provided are correlations between covariates (e.g., the correlation between age and nonverbal IQ is .109 in children with ASD). *Cor* refers to the overall correlation across groups.

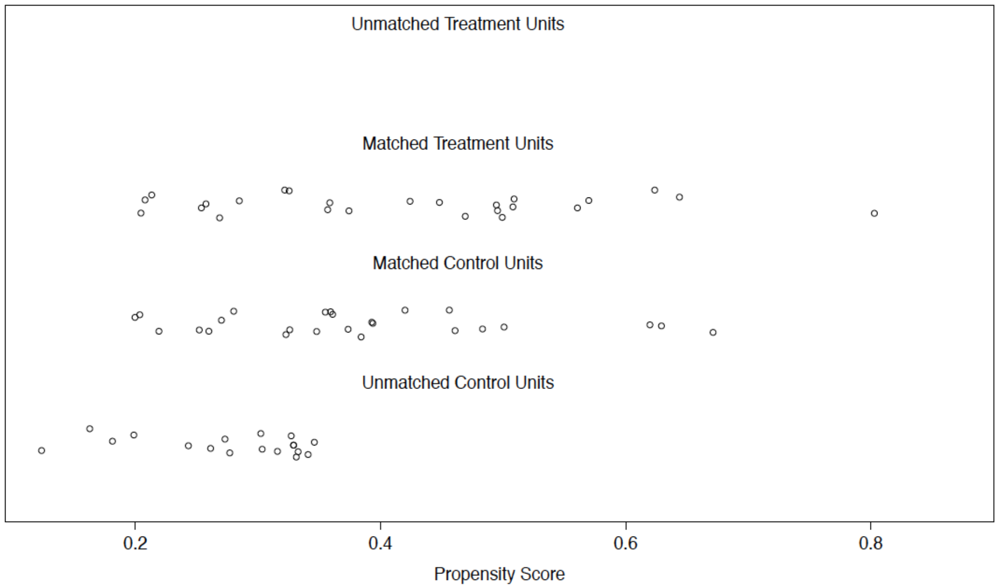
Next, I used the MatchIt package (Ho et al., 2011) in R version 3.3.0 (R Core Team, 2016), which calculates a propensity score for each participant and then matches groups based on those propensity scores. MatchIt includes multiple methods to match participants based on their propensity scores, and further detail on the different methods can be seen in Ho et al. (2011) and Stuart (2010). I used the *nearest neighbor* and *optimal* matching methods, and then evaluated whether groups were well matched using either method. To use MatchIt, you specify your dataset (i.e., each participant and their scores on the covariates), select the method you want to use (e.g., nearest neighbor or optimal), specify the covariates to include in the propensity score (e.g., IQ, age), and specify how you want to calculate the score itself (default is logistic regression). Then, after you run your command, MatchIt provides a dataset that includes the children in the matched group that were selected by the method[5].

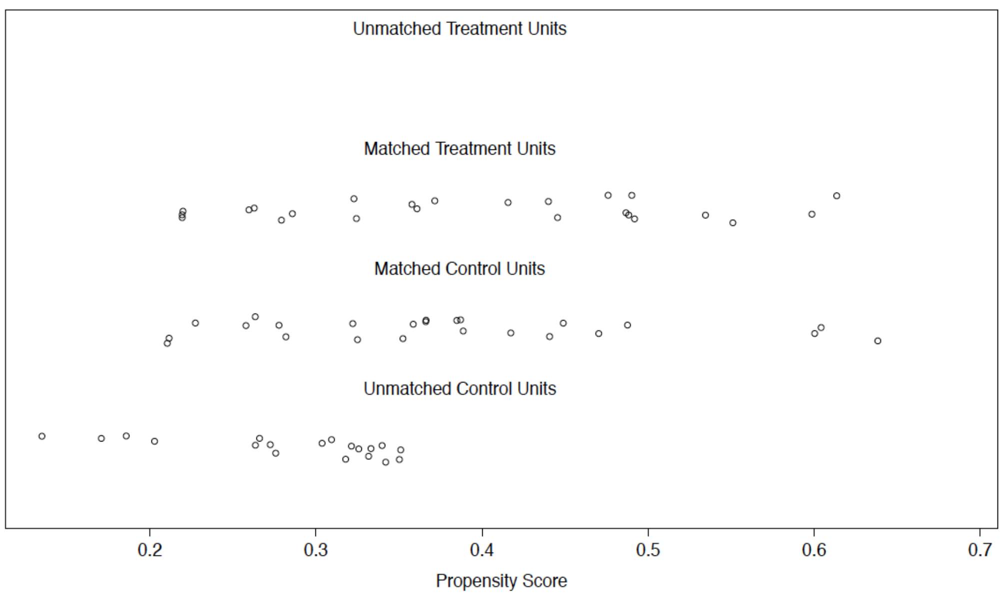
Nearest neighbor and optimal matching methods were first conducted with 25 children with ASD and 43 TD children. Although these methods use different algorithms, the same participants were ultimately selected with both methods using these samples. However, as seen below in Figure A 2, visual inspection of the propensity score distributions revealed one outlier in the ASD group, suggesting that matching could be improved by removing this outlier. Therefore, matching was conducted again with nearest neighbor and optimal matching, after removing this outlier in the ASD group. This time with 24 children with ASD and 43 TD children, nearest neighbor and optimal methods resulted in slightly different samples. By this

[5] *Exact* and *subclass* matching were not possible in MatchIt with this sample. *Ratio* matching was not used because there were not enough TD children (n = 43) for matching beyond a 1 to 1 match with the full ASD sample (n = 25). *Full* matching was not used because this requires using a weighted propensity score in a regression model, which can require more advanced knowledge to diagnose appropriately (King & Nielsen, 2016).

point, I have three different options of matched groups: the same sample determined from nearest neighbor and optimal matching with 25 children with ASD and 25 TD children, and two samples that include 24 children with ASD but slightly different groups of 24 TD children determined from nearest neighbor and optimal matching methods.

Figure A 2. Propensity Score Distributions Before (top) and After (below) Removing Outlier





These figures were created using MatchIt. Treatment units = children with ASD, Control Units = TD children. The figure on the top depicts an outlier in the ASD group with a propensity score of .8, indicating an 80% probability of being designated to the ASD group given the covariates included in the score, which in this case is nonverbal IQ and age. This score of .8 was seen as high given that there were no TD children who had propensity scores in this range. The figure on the bottom indicates that removing this participant resulted in more similar overlapping distributions with no clear outliers.

After a sample of matched children has been selected from the matching method, the next step is to evaluate, in each sample, how well groups are matched on the propensity scores themselves, as well as each covariate that was included in the propensity score (Blackford, 2007; Ho, Imai, King, & Stuart, 2007; Stuart, 2010). The strength of matching on propensity scores were determined from recommendations by (Rubin, 2001). Propensity scores were considered well matched between groups when the standardized mean difference of propensity scores was less than 0.25 (using the standard deviation of the ASD group as the denominator) and the variance ratio was close to 1 (i.e., within 0.5 to 2). Although both matching methods resulted in improved distributions based on visual inspection and numerical values as seen in Table A 1, the optimal method with 24 children with ASD and 24 TD children resulted in the smallest standardized mean difference and a variance ratio closest to 1.

Guidelines to evaluate well matched groups on each covariate (e.g., nonverbal IQ, age) included Cohen's *d* close to 0, variance ratios close to 1, and *p* values > .5 (Kover & Atwood, 2013; Mervis & Robinson, 1999; Mervis & Klein-Tasman, 2004). The use of descriptive statistics of Cohen's *d* and variance ratios are recommended as alternatives to inferential statistics such as *p* values, due to difficulties with establishing equivalence with inferential statistics (Kover & Atwood, 2013). Continuous covariates were also examined visually with quantile-quantile plots (Q-Q plots) and boxplots. Cohen's *d*, variance ratios, and *t*-tests for nonverbal IQ and age demonstrated that groups were best matched (relative to the other samples) with the sample of 24 children with ASD and 24 TD children determined with the optimal method.

Lastly, before determining the final matched sample, covariates not included in the propensity score (Stuart, 2010) and the distribution of children for randomized factors (i.e., cue condition order, context presentation order) were verified (King & Nielsen, 2016). Examination of these factors did not result in any reason to change the matched groups, thus the optimal matching method was used to determine our final matched sample from 24 ASD and 24 TD children.

Table A 1. Standardized Mean Differences, Cohen's *d* and Variance ratios (vr) for Nonverbal IQ, Age and Propensity scores

	25 ASD, 43 TD		25 ASD, 25 TD		24 ASD, 24 TD		**24 ASD, 24 TD**	
	Full Sample		Matched sample: both methods		Matched sample: nearest neighbor		**Matched sample: optimal**	
	d	vr	*d*	vr	*d*	vr	***d***	**vr**
Nonverbal IQ	-.53	.92	-.15	1.13	-.14	.99	**-.09**	**.91**
Age	.15	1.38	.27	1.61	.22	1.36	**.11**	**1.27**
Propensity Score[a]	.53	1.70	.24	1.46	.24	1.32	**.14**	**1.04**

[a] In accordance with Rubin (2001), standardized mean differences of propensity scores were calculated with the standard deviation of the treatment group, in this case the ASD group, as the denominator.

The matched sample with both methods refers to both the nearest neighbor and optimal method. The standardized mean difference scores for nonverbal IQ and age were calculated using a pooled standard deviation in the denominator using Cohen's *d* calculations (Kover & Atwood, 2013). Positive Cohen's *d* values indicate higher values in ASD children. Variance ratios above 1 indicate larger variances in ASD children. This table exhibits that although relative to the full sample, all methods indicated improved covariate balancing between groups, the optimal matching method (in bold) with 24 children with ASD and 24 TD children resulted in the smallest standardized mean difference and variance ratio between groups on propensity scores and the individual covariates.

Appendix B. Stimuli Creation and Pilot Testing for Word Learning Objects

Novel Labels

Novel labels were created in a prior norming study in our lab (Howarth, 2010). Eighteen labels were created using a nonword generator, WordGen (Duyck, Desmet, Verbeke, & Brysbaert, 2004), which measured the degree to which two-letter combinations were common in English and French. These scores were accumulated for each pair of letters in the word to result in an overall score for each label in English and in French. The 18 bi-syllabic novel labels were chosen by Howarth (2010) because they received a score of being "moderately common" in both languages by WordGen. Next, 10 native English and 10 native Quebec-French speakers rated how plausible the label sounded in their respective language on a score of 1 to 5, with 1 as very plausible and 5 as not plausible.

A subset of four novel labels were chosen for the present study. These novel labels were chosen based on the following criteria: 1) started with a different sound, 2) had a similar rating in English and French, and 3) had not been used in a prior word learning study in our lab given that some children would be returning participants. Plausibility ratings can be seen in Table B 1 and international phonetic alphabet spelling and stress can be seen in Table B 2. Novel labels and phrases (e.g., *pagoune*, *Where is the pagoune*, *Now point to the pagoune*) were recorded in English and French by a bilingual female English Quebec-French speaker in a sound proof booth using a Marantz PMD660 recorder. Final versions of recordings were selected in collaboration with three bilingual English Quebec-French speakers, who selected the best recording based on what they felt was well-paced, a balance between sounding the most enunciated, not too over or under exaggerated (to limit similarities to what may be perceived as child-directed speech, also proposed to be an intentional cue; Csibra and Gergely (2009)). Recordings were trimmed in

Praat (version 5.3.44) and the intensity modified to a mean of 75dB. Table B 3 exhibits the durations of the novel labels and phrases in English and French. Each target object had its own novel label. Please see the novel objects section below regarding the naming of novel objects.

Table B 1. Plausibility Ratings for Novel labels in English and French

	English	French
fopam	2.7	3.4
mimole	3.5	3.1
nalip	3.1	3.4
pagoune	3.6	3.6

Table B 2. International Phonetic Alphabet Spelling of Target Labels and Stress

	English	French
fopam	**f o** p ə m	f o **p a m**
mimole	**m i** m ɔ l	m i **m ɔ l**
nalip	**n æ** l ɪ p	n a **l i p**
pagoune	p ə **g u n**	p a **g u n**

The bold indicates the syllable that is stressed.

Table B 3. Durations (s) of Novel Labels and Phrases in English and French

	Isolated (spoken twice in this duration)		*Where is the _______?/ Ou est le _______?*		*Now point to the ______?/ Maintenant point le _____?*	
	English (gap)	French (gap)	English	French	English	French
fopam	1.479 (.348)	1.537 (.390)	1.068	.963	1.282	1.533
mimole	1.468 (.394)	1.507 (.312)	1.114	.940	1.238	1.473
nalip	1.488 (.381)	1.460 (.334)	1.095	.929	1.349	1.526
pagoune	1.556 (.445)	1.428 (.339)	1.009	.934	1.280	1.470

Isolated labels were spoken twice, and the duration of the gap between labels are presented in parentheses.

Novel Objects

Four target objects and four distractor objects were designed for this study. Target objects were designed to be visually less interesting than distractor objects by having limited features and decorations such as only two colors per target, and simple shapes and materials (e.g., triangle, paper). These limited features were also thought to make the target objects easier to describe for children this age, which was done in consideration of the word description measure. Distractor objects were designed to be perceptually salient by being shiny, multi-colored, and more detailed than target objects (e.g., multiple little beads, gemstones were included on distractor objects but not on target objects). Each target and distractor object had a unique cause and effect function and all objects were roughly similar in size. Careful consideration was taken to create unique objects such that children would be able to demonstrate learning of specific target objects during the in-depth learning measures. Objects were piloted with 10 English and 10 Quebecois-French adult speakers to verify how interesting the objects were (object ratings), and the labeling of the objects (label ratings). Additionally, adults were asked to provide descriptions of the target objects to examine how specific each description would be for each of the target objects. Because of time and recruitment constraints, pilot testing was completed with adults instead of children. All adults provided consent prior to their participation and the testing session lasted approximately 30 minutes; adults volunteered to participate and were not given compensation.

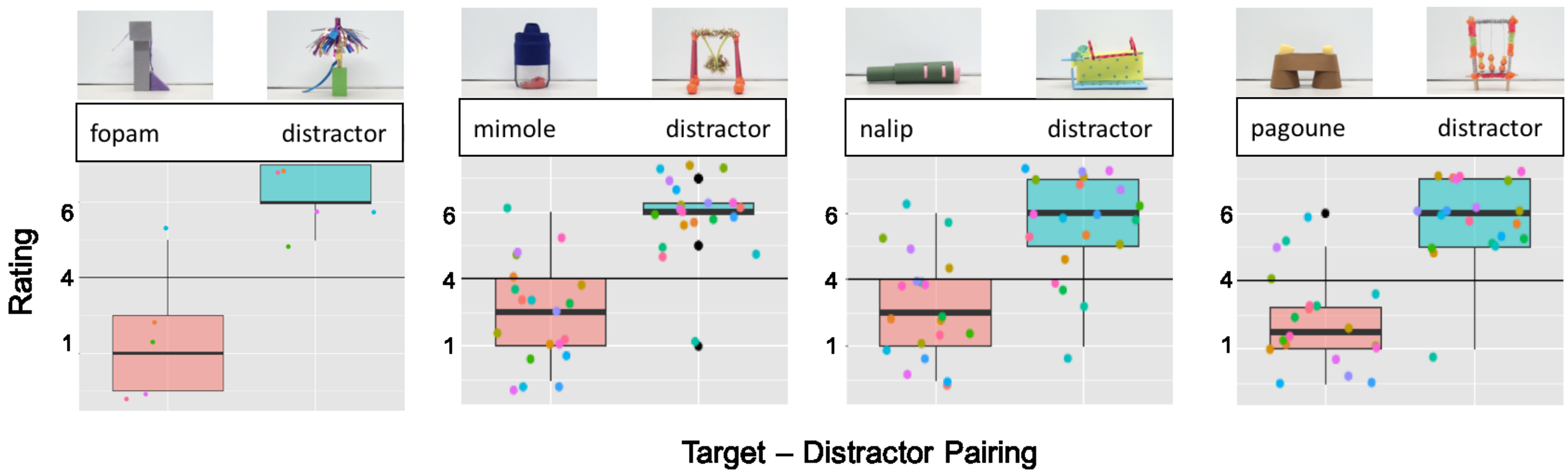
For object ratings, adults were shown an image with a target object on one side paired with a distractor object on the other side (the actor did not appear). Adults were asked to rate how interesting each object was on a scale of 1 to 7, where a score of 1 was not interesting, 4 was neutral, and 7 was very interesting. They were asked to give their first reaction, without thinking too long about their response and were shown a total of 16 images, which resulted in all possible target and distractor object combinations (4 targets x 4 distractors = 16 combinations). The side of target and distractor objects was counterbalanced across adults and the order of images was randomized.

Next, adults provided label ratings where they were shown a target object and listened to one of the four novel labels and asked to rate how well the label fit the object. A score of 1 was that the label did not fit the object at all, 4 was neutral, and 7 was that the label fit the object well. Each of the four novel labels were played with each target object, resulting in adults rating 16 label-target object pairings. Order of label-target pairings were randomized.

Finally, adults described familiar and target objects. Familiar objects were shown only as images because their function was never shown in the video, and target objects were shown with their function. Adults described the familiar objects first so that they could practice with known objects before describing unknown target objects. Adult descriptions were examined to prepare for what to potentially expect from children's descriptions, and were used to begin thinking about how to code semantic features; these findings will not be discussed further. For images and descriptions of the object functions, please refer to Figure B 3.

The following results examine adult pilot ratings for objects (how interesting the target and distractor objects were relative to each other), and labels (how well the label fit the target object). Because there did not appear to be language effects from visual inspection of the data, results from pilot testing were collapsed across English and French speakers. Findings from object ratings indicated that for all 16 target-distractor pairs there was an approximately 3-point median difference in ratings between target and distractor objects, with some variation for different pairings (target object *M* range 2.55 – 3.6, *Mdn* range 2.5 – 3.5; distractor object *M* range 5.45 – 6.1, *Mdn* range 6 – 6.5). These results confirm that adult participants found target objects to be less interesting than distractor objects. To minimize counterbalancing options and reduce variation in pairings, four fixed target-distractor pairs were used in the word learning study. These fixed pairs were chosen because they were separated by a minimum median difference of 3 points, and objects were similar in height and width, such that objects that were taller were paired with each other, and objects that were shorter and longer were paired with each other. Only one target object was slightly revised based on object ratings (i.e., changing one object from having a triangle top to a square) and re-evaluated by adults. No differences were seen between target objects on their ratings after this revision, thus the revised object was kept and no additional changes were made. Figure B 1 depicts the object ratings for the fixed target-distractor pairings.

Figure B 1. Ratings of the Final Fixed Target and Distractor Pairs by Adult Participants

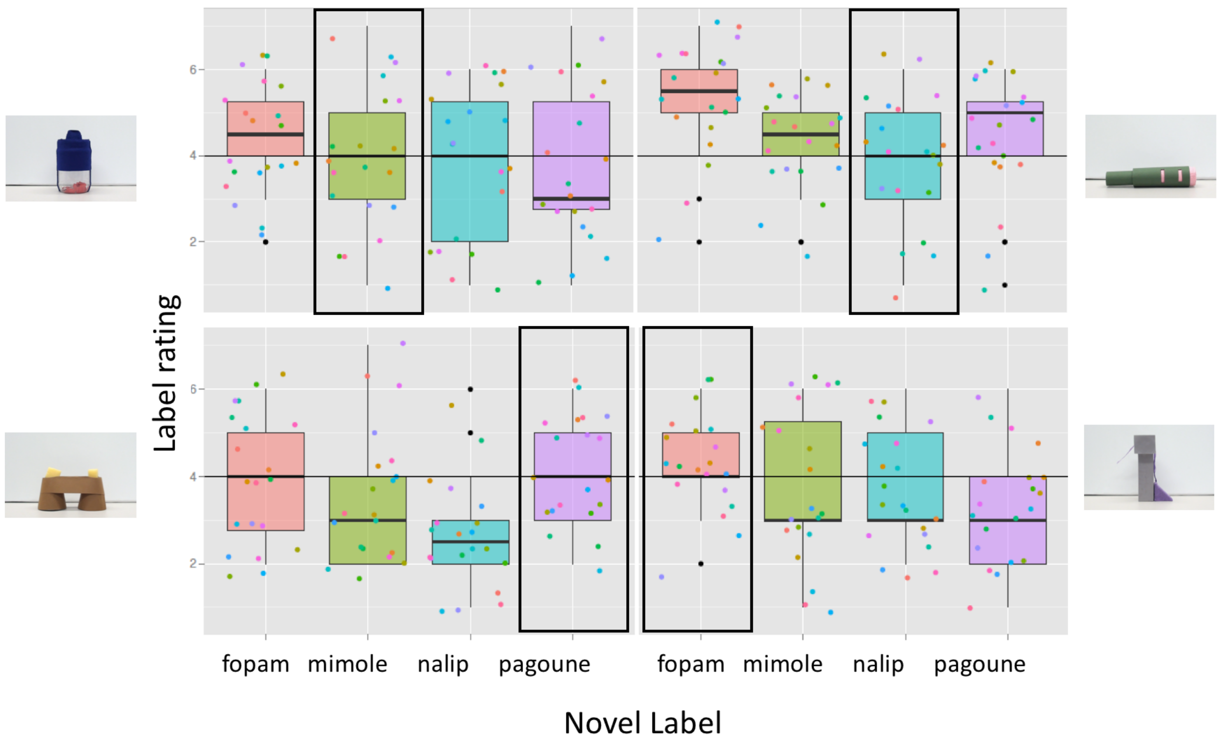


Each box represents the interquartile range, or the middle 50% of the data. The whiskers on either side of the boxplot represent the bottom 25% and the top 25% outside of the interquartile range. Dark black points are outliers that are greater than 1.5 times the interquartile range. The dark black bar in each box is the median rating for that label. Colored points represent ratings of individual participants. The black horizontal line was included to mark the neutral rating of 4.

Images of the target (fopam, mimole, nalip, pagoune) and their respective distractor objects are presented above the boxplots of their ratings. The y axis ranges from 1 to 7, where a rating of 1 means that the object was not interesting relative to the other object and 7 means that object was very interesting relative to the other object. The fopam-distractor pairing depicts fewer points because data are from the revised pilot testing where only a quarter of the participants rated this particular pairing. This pilot data demonstrates that all four pairings were separated by a minimum median difference of 3 points, which confirms that the target object was perceived to be less interesting than the distractor objects for adult participants (higher boxplot).

Findings from label ratings revealed that most label-object pairings were rated within the range of 2.5 to 5.5 points. Therefore, adults were generally neutral to the pairing of any novel label with any target object. However, because there was slight variation for certain novel labels as seen in the range above, to account for these differences, one novel label was fixed to a specific target object as seen above with target-distractor pairings. It was important that all pairings between labels and target objects had a similar score so that opportunities for word learning was similar across objects. From visual inspection of the pilot data, this was best achieved by selecting pairings where the median score was always a neutral score of 4. Figure B 2 depicts the label ratings for the label-target object distractor pairings.

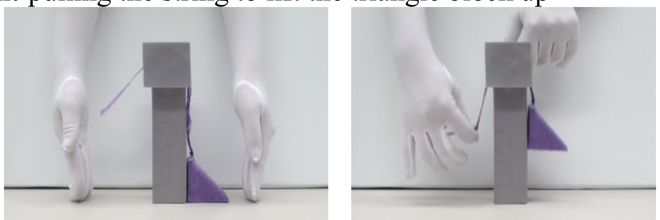
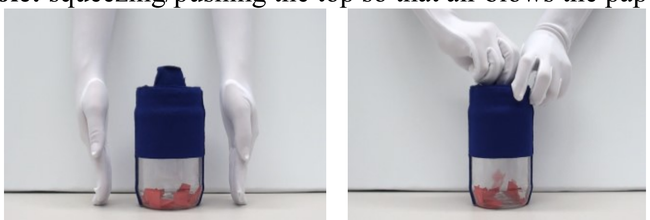
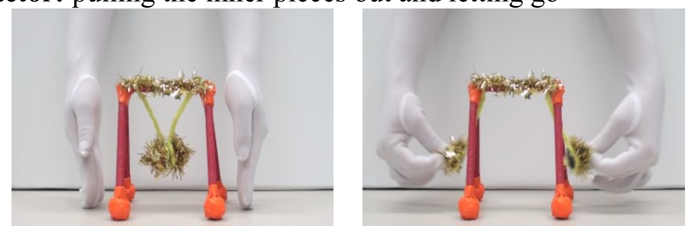
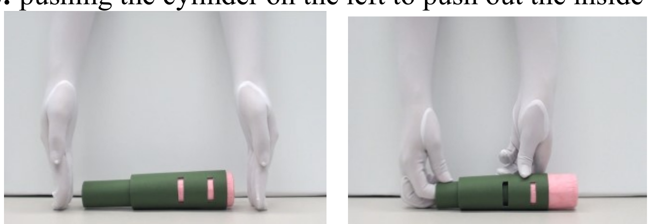
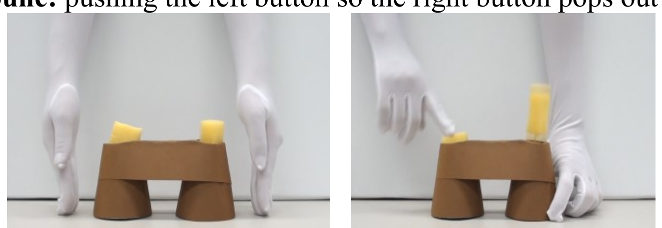
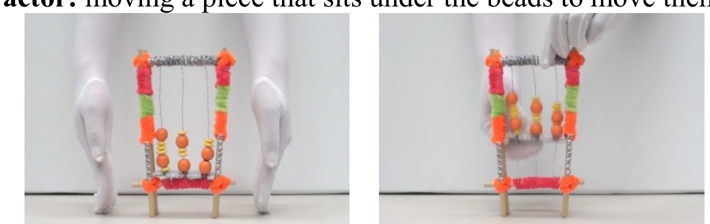
Figure B 2. Ratings of the Final Fixed Target and Label Pairs by Adult Participants



Each box represents the interquartile range, or the middle 50% of the data. The whiskers on either side of the boxplot represent the bottom 25% and the top 25% outside of the interquartile range. Dark black points are outliers that are greater than 1.5 times the interquartile range. The dark black bar in each box is the median rating for that label. Colored points represent ratings of individual participants. The black horizontal line was included to mark the neutral rating of 4.

The y axis ranges from 1 to 7, where a rating of 1 means that the label did not fit the target object and 7 means that the label fit the target object well. The x axis includes the four possible novel labels (fopam, mimole, nalip, pagoune) that were presented with each target object. Images of the target objects paired with the target label (in the black outline) are shown next to each set of graphs. The boxplot outlined in black is the label that was chosen for that target object, which was selected because it had a median of four and the interquartile range was symmetrical around the median (except for the label of fopam). This pilot data demonstrates that the final pairings all had the same median rating of 4, indicating that the labels were similar in how well they fit their respective target object.

Figure B 3. Images and Descriptions of Target and Distractor Object Functions

Target Object and its Function	Distractor Object and its Function
fopam: pulling the string to lift the triangle block up 	**distractor:** pulling the lower string to twist the loose strips on top 
mimole: squeezing/pushing the top so that air blows the paper inside 	**distractor:** pulling the inner pieces out and letting go 
nalip: pushing the cylinder on the left to push out the inside cylinder 	**distractor:** moving the rectangle slide down so the ball rolls 
pagoune: pushing the left button so the right button pops out 	**distractor:** moving a piece that sits under the beads to move them up 

Each function had a simple cause and effect that started and ended with the hands and objects in the same stationary rest position: hands on either side of the object. The hands began moving from rest at the start of the 4.3 s of the cue portion in the teaching phase (from the start of the cue shift to the target object) and came back to the same rest position before the end of the 4.3 s of the cue portion (by the end of the cue shift back to direct gaze/static circle).

Appendix C. Stimuli Creation and Pilot Testing for Action Prediction Blocks

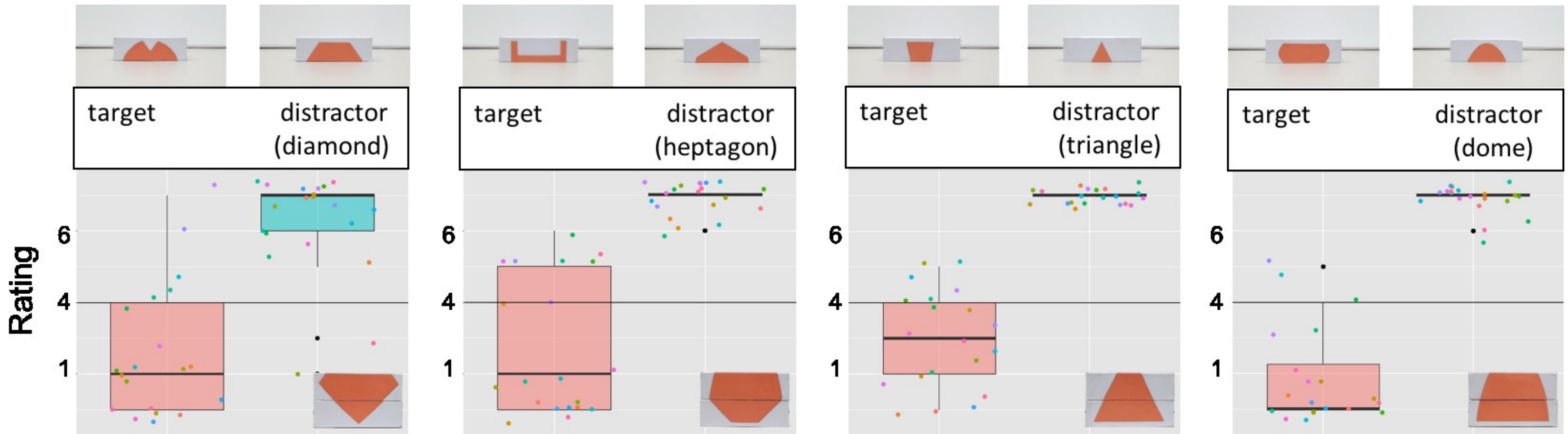
Four experimental block sets were designed for this study. Each set included four blocks with two of the blocks forming the base of a pre-stacked tower. The other two blocks were two different options to complete the tower. One of the blocks was a distractor, or the block that participants would more likely gravitate to, and completed a conventional shape. The other block was the target, or the block that the cue would be directed to, and completed an unconventional shape.

Conventional shapes were intended to be simple so that children would quickly recognize the shapes and discern the block to complete the conventional shape. Conventional shapes included a triangle, diamond, pentagon, and a heptagon, which were selected because they were vertically symmetrical. Unconventional blocks were designed such that the image on the block that completed the unconventional shape had a similar number of sides as the image on the block that completed the conventional shape, be vertically symmetrical, and importantly, that images on both target and distractor blocks could match the image on the pre-stacked tower to complete the shape along the same points (so that the unconventional block was still a valid option to complete the shape).

The shapes created for this study were piloted with the same 10 English and 10 Quebecois-French adult speakers who rated the word learning objects and provided consent prior to their participation. This testing session lasted approximately 10 minutes. Adults volunteered to participate and were not given compensation. Adults were shown an image with the target block on one side, the distractor block on the other side, and the pre-stacked tower in the center. The actor in the video was not shown in the images during pilot testing. Adults were asked to rate the target and the distractor block for how well the block completed the shape on a scale of 1 to 7. A score of 1 was that the block did not complete the shape well, 4 was neutral and 7 was that the block did complete the shape well. Adults were asked to give their first reaction, without thinking too long about their response. The side presentation of the target and distractor blocks were counterbalanced, and the order of sets was randomized across participants.

The following results examine adult pilot ratings for how well either the target or distractor block completed the shape. As in word learning pilot testing, results were collapsed across English and French speakers. Findings from pilot testing indicated that the target and distractor blocks were separated by a minimum score of 4 points for the dome and triangle. However, for the heptagon and diamond the differences were smaller at 1 point and 3 points, respectively. Therefore, the unconventional blocks for the heptagon and diamond were revised and re-evaluated by the same adults. The revised ratings resulted in a minimum 4-point difference between the target and distractor blocks for all four sets. Figure C 1 depicts the shape and ratings for the four experimental block sets used in the study.

Figure C 1. Ratings of the Target (Unconventional Shape) and Distractor Blocks (Conventional Shape) by Adult Participants



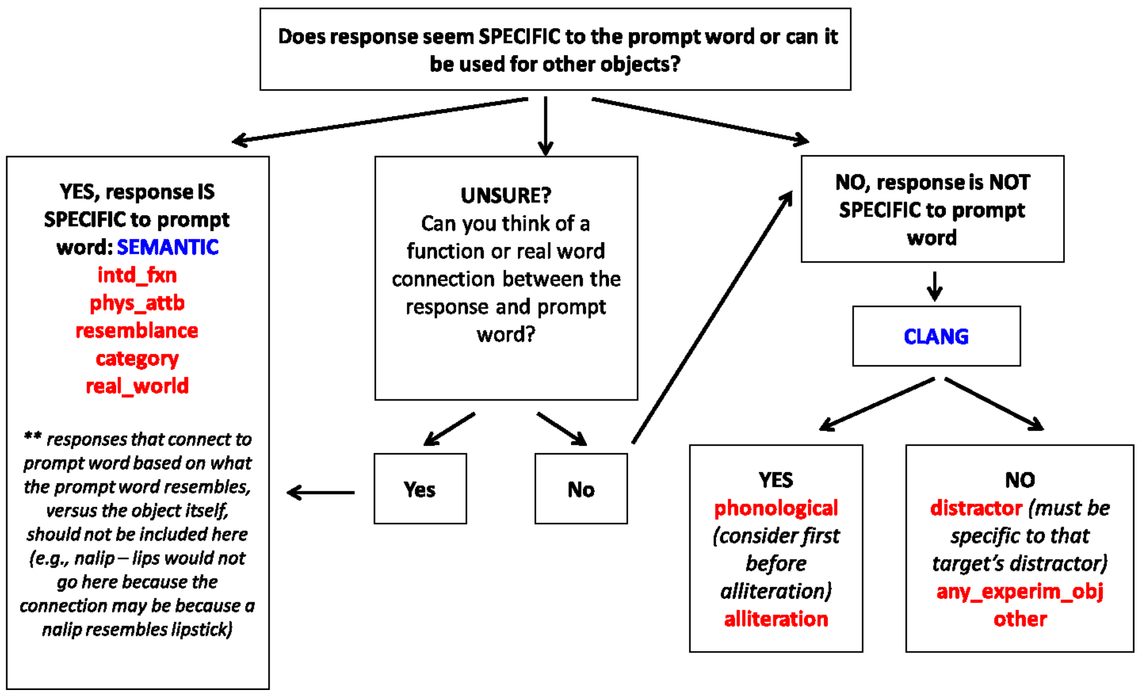
Each box represents the interquartile range, or the middle 50% of the data. The whiskers on either side of the boxplot represent the bottom 25% and the top 25% outside of the interquartile range. Dark black points are outliers that are greater than 1.5 times the interquartile range. The dark black bar in each box is the median rating for that label. Colored points represent ratings of individual participants. The black horizontal line was included to mark the neutral rating of 4.

The pre-stacked tower is seen in the bottom right corner, and images of the target and distractor blocks (as they were presented in the video) are seen above their ratings. Target blocks create an unconventional shape while distractor blocks create a conventional or familiar shape (name of the conventional shape is seen in parentheses). The y axis ranges from 1 to 7, where a score of 1 rates the block as not completing the shape well and 7 as completing the shape well. The black horizontal line represents a neutral score of 4. This pilot data demonstrates that the final four sets had target-distractor block pairings that were separated by a minimum difference of 4 points. This difference confirms that the target block was perceived to be a less acceptable fit for the shape, and that the distractor block was confirmed as a more acceptable choice by adult participants.

Appendix D. Decision Tree for Coding Word Associations

used for target objects although can be used for familiar objects

NR/inflections/repetitions = CODED UNDER ERROR



This decision tree was developed in collaboration with the coder. I included this decision tree to demonstrate that word association coding was operationalized, thus it was not made for general use although it can serve as a reference. A manual for word association coding is available upon request. All words in red represent the different coding groups possible to a coder. The font in blue (i.e., SEMANTIC, CLANG) was done to serve as visual markers for the coder. The shorthand above for coding groups was used in communication with the coder.

The six coding groups in this dissertation were ***semantic***, ***clang***, ***distractor, any experimental object***, ***error***, and ***other***. ***Semantic*** was initially made up of subgroups of *intended function* (intd_fxn), *physical attributes* (phys_attb), *resemblance*, *category*, and *real world knowledge* (real_world), though after further consideration when coding was complete, semantic responses for target objects excluded the coding groups of *category* and *real world knowledge* (see Methods section for explanation). ***Clang*** included subgroups of *phonological* (shared similar sounds), and *alliteration* (only shared the first letter). ***Errors*** included when children did not provide a response (*NR*) or said *I don't know*, *inflections*, or *repetitions*. Coding groups of ***distractor***, ***any experimental object (any_experim_obj)***, and ***other*** were left on their own, because it was not clear how else to summarize these coding groups. For further definitions and examples of each category, please refer to the Chapter 1 Methods section.

Appendix E. Decision Tree for Object Identification in Word Description

1. When colors, functions, and other physical attributes seem to reference target_object (T)

other physical attributes only

1. Any reference to shapes that make it specific?
- If yes = T
- If no = see 2 below

2. Any other physical attributes that make it specific?
- If yes = T
- If no = other-general

color & no function

2 colors = T

1 color that CANNOT be confused with distractor = target_object
(even if child provides other colors that are incorrect)

BUT if v2 then consider gentarget option if color refers to an eprime image

1 color CAN be confused with competitor?
(even if child provides additional colors that are incorrect)

1. Any reference to shapes that make it specific?
- If yes = T
- If no = see 2 below

2. Any other physical attributes that make it specific?
- If yes = T
- If no = other-general

function & no color or "any color"

2 functions = T

1 function?
1. Any reference to shapes that make it specific?
- If yes = T
- If no = see 2 below

2. Any other physical attributes that make it specific?
- If yes = T
- If no = other-general

color & function

minimum:
1 color & 1 function = T
(even if child provides additional colors that are incorrect)

BUT if v2 then consider gentarget option if color refers to an eprime image

2. When colors, functions, and other physical attributes seem to reference any other object, use the same schema as above but the final decision of "T" would be replaced with one of the following:
- distractor
- diffdistractor-name
- difftarget-name
- other-nameor other-both when description fits both target and distractor

3. Other possible identifications
- gentarget-name (this is ONLY for when child describes the generalizations from the eprime game)
- dont_know

NOTE: if same set of descriptions used repetitively, code as other-general

This decision tree was developed in collaboration with the coder to operationalize object identification coding, thus it was not made for general use although it can serve as a reference. The shorthand above for coding groups was used in communications with the coder (e.g., v2 = visit 2, eprime image = word generalization task).

Appendix F. Building Linear Mixed Models

One of the benefits of mixed models is the ability to include both random and fixed effects. Fixed effects in our models included cue condition (gaze versus arrow) and group (ASD versus TD). For some measures, an additional fixed effect represented performance over multiple time points (e.g., multiple video phases). All main effects and interactions of fixed effects were included in the models. In addition to fixed effects of interest and random effects, mixed models can include other continuous fixed effects, such as covariates of children's IQ or age. No covariates were included in the models because groups were already well matched on nonverbal IQ and age, and small sample sizes limited including continuous fixed effects as well as the fixed effects of interest (categorical variables of cue condition, group, and time point). I only included a random intercept of participant, without any slopes (i.e., how each participant varies on each level of the respective fixed effects). Though there are many approaches to specify more complicated random effect structures, the process below details why the most parsimonious model including only a random intercept of participant, was chosen for the current dataset.

Specification of random intercepts and slopes is a complex process and can be guided by how observations are clustered. For example, in the current study, participants provided repeated observations (up to 4 observations, 2 per cue condition). Therefore, including a random intercept of participant would capture the variation in each participant's overall mean on the dependent variable. By including a random intercept of participant in the model, the findings of our fixed effects have accounted for each participant's individual variation, thus we can better generalize findings of our fixed effects beyond aspects our specific sample of participants.

In addition to including random intercepts, we could also take into account random slopes. For example, fitting a random slope of cue condition by participant will account for the fact that the magnitude and direction of the difference between gaze and arrow conditions on a particular measure can vary by participant. If there is still a significant fixed effect of cue condition even after accounting for this variation, then this difference between gaze and arrow conditions is more likely to be generalizable beyond the sample in this dissertation. Consequently, including random slopes in the model is thought to be critical to interpret within-subject fixed effects because including random slopes avoids Type I errors, and meets assumptions of conditional independence for a mixed model (Barr, Levy, Scheepers, & Tily, 2013).

There are two opposing suggestions for specifying random effects structures. On the one hand, Barr and colleagues (2013) recommend trying to fit a maximal random effect structure which includes all random intercepts, slopes, and slope interactions, except in cases where there is only 1 observation per level. With so few observations, random slopes should not be used because the slope variance would be confounded with trial-level error, thus a random intercept only model is best. On the other hand, Bates and colleagues (2015a) suggest that parsimonious models, or a model as simple as possible, are sufficient, particularly when the main interest is to examine factorial contrasts. For our measures, participants in each group mostly contributed 2 observations per level of cue condition, although on some measures 1 observation per level can also be seen (e.g., the number of semantic features for word description). When main effects were collapsed across time points, then for each participant there were a possible 2 - 6 observations per level of cue condition depending on the number of time points. Given that at times there was a low number of observations (Barr et al., 2013), and my main interest was factorial contrasts (Bates et al., 2015a), I determined that for this dissertation, a parsimonious model included a random intercept of participant and fixed effects of interest (e.g., cue condition, group, and time point where applicable).

However, because the majority of the measures included more than 1 observation, albeit at times only 2 observations per level within each time point, and Barr et al. (2013) stated that slopes are critical to interpret within-subject effects such as cue condition, for each measure a maximal random effect structure was evaluated against a parsimonious model (i.e., random intercept of participant only). Therefore, maximal random effect structures included all slopes and interactions for within-subject fixed effects (cue condition and video phase) by a random intercept of participant, a random intercept for between-subject effects (group), and all correlation terms. In cases where the models did not converge or if the correlation parameters of the random effects were too high (close to ±1 in Barr et al. 2013, but defined as > ± .90 in this dissertation), the maximal model was simplified by first removing correlations. If models did not converge or correlations did not improve, then slope interactions were removed, followed by the slopes themselves.

The comparisons between maximal (or as maximal as possible) and parsimonious models revealed that, as expected due to few repeated observations, random slopes often did not add explanatory value for most models. This lack of explanatory value was determined by 1) higher AIC values in maximal models (higher AIC values indicate a worse fit), and 2) non-significant or significantly worse goodness of fit for maximal versus parsimonious models. These comparisons confirmed that parsimonious models were the best fit to interpret our fixed effects for most models. However, for some measures (e.g., word learning eye tracking diagnostics and the word generalization response time data), the maximal model was used because including random slopes resulted in a better fit (lower AIC, significantly improved goodness of fit).

Appendix G. Data Cleaning Protocol and Results

Exclusion of Children from Eye Tracking Data

Two children with ASD from the matched group were not included in the eye tracking analyses because 1 had strabismus and the other was not able to be recognized by the eye tracker. In the full TD group, 2 children were removed from eye tracking analyses because 1 had a history of strabismus (surgically corrected), and the other could not calibrate according to my criteria (see below); however, these 2 children were not selected as part of the matched group using MatchIt (Ho et al., 2011). This left 22 children with ASD and 24 TD children possible in the matched group for eye tracking analyses.

Calibration

One of the first steps to validate eye tracking data begins with how participants are calibrated prior to watching the videos. Calibration is done by showing children points (usually a minimum of 5 points) that appear on the screen, and asking participants to look at the points. When participants are looking at these points, the eye tracker is collecting information on children's accuracy and precision of looking at the point on the screen. FaceLAB provides a mean angular error score for each eye after calibration, which is a value that represents the accuracy and precision of the movements of their measurement. Using a visual angle calculation, which considers the child's distance from the screen, a mean angular error score of 5 was determined as an acceptable cut-off for the current study. This mean angular error score ensured that all children were held to the same standard of accuracy and precision that least compromised the integrity of the data for the main research questions (e.g., data points that were in the location of the area of gaze could not be mistaken for the location of the area of the arrow).

I used a nine-point calibration, with small moving animals to help keep children attentive to the screen. After children watched the nine-point calibration, the mean angular error was immediately shown for each eye. The mean angular error is the mean of multiple sampling points, which is a value that represents the quality of calibration of the points that were shown on the screen. To determine the maximum mean angular error valued allowed in this study, I used a visual angle calculation which includes a mean angular error value and children's distance from the screen. A value of 5 was deemed as an acceptable cut-off for this study. This mean angular error value allowed for a potential offset radius of 4.06 cm, meaning that when children were looking at a specific point on the screen that they could be off by 4.06 cm from this point. This distance did not overlap between any areas of interest that were directly compared in this study (e.g., areas of the gaze versus the arrow = 13 cm apart from their closest points, areas of the target and distractor object = 22 cm apart from their closest points, areas of the cue to the distractor or target object = 4.8cm apart from their closest points).

I attempted to calibrate children before the start of each block since there was a short break between each block. However, there were children in both groups that couldn't be calibrated before watching the block 2 videos, and in these cases I re-used the settings from block 1. It was unclear why in certain cases children were unable to be re-calibrated, though one reason may be due to eye fatigue nearer the end of the testing session. A similar number of children in each group needed to re-use the settings from block 1, because they could not re-calibrate both eyes under 5 degrees mean angular error before block 2 (8 ASD, 5 TD). Children's mean angular error scores were compared between groups for those in the word learning sample and those in the action prediction sample, because children slightly differed between samples (see Chapter 2 Methods). For children who were able to be calibrated in both blocks, mean angular error scores were compared within groups in block 1 and block 2 to examine if there were block effects in calibration for these children.

Word learning sample. In block 1, the mean angular error did not differ between children with ASD (*M* = 1.99, *SD* = .96) and TD children (*M* = 1.68, *SD* = .80), $t(41.02) = 1.23$, $p = .23$, or in block 2 between children with ASD (*M* = 1.92, *SD* = .79) and TD children who were able to re-calibrate (*M* = 1.48, *SD* = .47), $t(19.97) = 1.90$, $p = .08$. Additionally, for the 14 children with ASD and 18 TD children who could be calibrated in both block 1 and block 2, I conducted paired t-tests to compare the mean angular error between blocks, and there were no significant differences between blocks, $t(31) = .44$, $p = .67$. Therefore, children were well calibrated before the start of the videos, there were no differences between groups, and there was no difference between blocks, indicating that for children who were re-calibrated in block 2 there were no fatigue effects.

Action prediction sample. As seen in the word learning sample, in block 1, mean angular error did not differ between children with ASD (*M* = 1.96, *SD* = .99) and TD children in block 1 (*M* = 1.74, *SD* = .85), $t(37.91) = .78$, $p = .44$, or in block 2 between children with ASD (*M* = 1.59, *SD* = .62) and TD children (*M* = 1.94, *SD* = .81), $t(19.23) = 1.31$, $p = .21$. Additionally, for 12 children with ASD and 18 TD children who were able to be re-calibrated, paired t-tests indicated no significant differences within groups between block 1 and block 2, $t(29) = .10$, $p = .92$.

Data Cleaning of Videos with Low Looking Times

One important part of data cleaning entails decisions about how to account for overall low looking times to videos (Venker & Kover, 2015). Overall low looking times, assuming calibration criteria were met, can be due to excessive movement or other technical issues with the eye tracker itself, as well as low looking by the child. Eye tracking studies have varied widely in the duration looking time required to retain data, ranging from 10% to 50% (Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2009; Venker & Kover, 2015; Vivanti et al., 2011), and Venker and Kover (2015) propose a general cut-off of 50% to strike a balance between including valid data and maximizing statistical power (a lower cut-off may result in excluding more trials). However, a 50% cut-off was considered to be too conservative for the current study, when taking into consideration the duration of the videos.

Eye tracking data were collected for each video phase and ranged from 4 seconds to 4.3 seconds for each phase used in analyses: baseline, cue portion of the teaching phase, and test phases (see Figure 3). I adopted a less conservative cut-off of 25% looking time, because 25% looking is approximately 1 second (e.g., 25 % of a 4 second video), which is a comparable duration or longer than prior studies with varying percentages for cut-off points (Fletcher-Watson et al., 2009 – 500ms out of 3000ms, Venker et al., 2013 – 50% of a 1700ms window (850ms), Vivanti et al., 2011 10% of a 4s window (400ms)).

Based on recommendations by Venker and Kover (2015), I used the following steps to clean data at the level of trial (i.e., one video sequence), condition, and participant for low looking times. One video sequence included video phases of baseline, teaching (label and cue portions), center, and test phases, and each child had 4 video sequences (2 per cue condition). First, video phases were dropped when children did not look for at least 25% of the video phase.

Next, eye tracking data were cleaned at the level of the video sequence. Video sequences were considered complete for analyses when overall looking times were above 25% to baseline, cue portion of the teaching phase, and test phases. In other words, a complete video sequence for the purposes of analyses meant children were looking more than 25% at video phases that most required children's attention to make sense of what was happening in the video. This ensured that the data analyzed in each video sequence were under approximately the same sequential viewing progression (e.g., analyzing the test phase meant that the child had seen the baseline and the cue portion before the test phase). Finally, if a child did not contribute at least one trial in each cue condition, then that child was dropped from eye tracking analyses. After data were cleaned, I used a linear mixed model to examine children's overall looking time to the video phase. While for all research questions I adopted a unit of 100 ms as the unit of looking time, which is commonly used as an indicator of children's cognitive processing (Gredebäck et al., 2009; Oakes, 2012), I used a smaller unit of time for overall looking time. For overall looking time I used a unit of 33 ms, because I wanted to use a more fine-grained measure of children's looking time. The following paragraphs detail in the word learning and action prediction samples the number of video phases and sequences that were dropped based on the criteria set above, and an analysis of children's overall looking time to the video in baseline, cue portion of the teaching phase (referred to as the teaching phase), and test phases.

Word learning sample. Low looking times were only addressed in static video sequences, because function video sequences were not analyzed in the current study. Video phases with less than 25% looking included the following number of phases per group: baseline (ASD = 0, TD = 1), teaching (ASD = 5, TD = 2), and test (ASD = 4, TD = 6). Four video sequences were dropped for children with ASD (affected four different children) and six video sequences were dropped for TD children (affected four different children). Dropping video sequences resulted in one TD child who did not have data for both cue conditions and thus was removed from eye tracking analyses. Therefore, with the 2 children with ASD excluded from eye tracking analyses, the final data set used in eye tracking analyses included 22 children with ASD and 23 TD children from the matched group.

In the word learning context, I used a linear mixed model to examine children's overall looking time using the dependent variable of the proportion of overall looking time. The proportion of looking time was calculated by dividing the overall looking time (based on a unit of 33ms) relative to the overall duration of the respective phase; proportions ranged from 0 to 1. A maximal model was specified that included random slopes for cue condition and video phase with a random intercept of participant (no interaction between cue condition and video phase in the slope), random intercept of group, and fixed effects of cue condition (gaze, arrow), group (ASD, TD), and video phase (baseline, teaching, test). The proportion of overall looking time was transformed using a power transformation due to the negative skew in the data (many children looking for a high percentage of time). As seen in Table G 1, there were no significant main effects of cue condition, group, or interactions between cue condition, group or video phase, but only a significant main effect of video phase. Post hoc tests revealed that children spent a significantly smaller proportion of overall looking time during test ($PMM = .80$, $SE = .09$) in comparison to baseline, $t(49.07) = 6.37$, $p < .001$, *PMM* difference = .11, $d = .70$, and a significantly smaller proportion of overall looking time during test in comparison to teaching, $t(61.08) = 5.43$, $p < .001$, *PMM* difference = .10, $d = .66$. However, the mean proportion of overall looking time of 78% during the test phase still seemed like children were looking during the majority of the test phase (raw $M = .78$, $SD = .20$, range = .25 - 1). There was no difference between baseline (PMM = .91, SE = .08) and teaching phases (*PMM* = .90, *SE* = .08), $t(58.68) = .71$, $p = .76$, *PMM* = .01, $d = .05$. In summary, the eye tracking diagnostics indicated no effect of cue condition, group, or interactions between cue condition, group, and video phase, except for a smaller proportion of looking time by the test phase. Therefore, in the case of any cue condition or group differences, these results suggest that potential differences on specific areas of interest would not be due to any systematic differences in children's overall attention to video phases.

Table G 1. Word Learning: Main Effects and Interactions of Overall Proportion of Looking Time During Video Phases By Group and Cue Condition

Factor	*df*	*F*	*p*
Cue condition (C)	1, 46.57	.001	.97
Group (G)	1, 1.18	.12	.78
Video phase (V)	2, 49.74	22.62	<.001***
C x G	1, 46.57	3.22	.08
C x V	2, 336.51	1.95	.14
G x V	2, 49.74	1.28	.29
C x G x V	2, 336.51	2.23	.11

*** $p < .001$

Action prediction sample. Video phases with less than 25% looking time included the following number of phases per group: baseline (ASD = 2, TD = 0), teaching (ASD = 3, TD = 1) and test (ASD = 3, TD = 2). Six video sequences were dropped for children with ASD (affected 5 different children) and five video sequences were dropped for TD children (affected 3 different children). All children provided data for at least one complete video sequence for analyses in each cue condition, thus no children were dropped from eye tracking analyses. Therefore, excluding three children with ASD and one TD child who were not included in the action prediction sample (see Chapter 2 Methods section), and excluding the 2 children with ASD who were not included in eye tracking analyses, the final data set used in eye tracking analyses included 20 children with ASD and 23 TD children from the matched group.

In the action prediction context, I also used a linear mixed model to examine children's overall looking time using the dependent variable of the proportion of overall looking time; the proportion variable was defined in the same way above. The model was specified with a random intercept of participant, and fixed effects of cue condition (gaze, arrow), group (ASD, TD), and video phase (baseline, teaching, test). The proportion of looking time was transformed using a power transformation due to the negative skew in the data (many children looking for a high percentage of time). As seen in Table G 2, there was a significant main effect of video phase, and significant two-way interactions of cue condition and group, and a two-way interaction of cue condition and video phase (ps < .05). The three-way interaction of cue condition, group, and video phase was marginally significant at $p = .05$. Yet because this three-way interaction just missed significance at our alpha < .05, I reviewed the three-way interaction. I did not perform post-hoc tests of the two-interactions because the three-way interactions would provide further information of how the two-way interactions vary by the third fixed effect.

The following 24 comparisons were included in the post hoc tests: 1) at each level of cue condition, group comparisons during each video phase (e.g., arrow condition: ASD during baseline vs. TD during baseline, 3 comparisons per group for 6 total), 2) at each level of group, comparisons between video phases for each cue condition (e.g., ASD: baseline phase in arrow condition versus teaching phase in arrow condition, 6 comparisons per group, for 12 total), and 3) at each level of video phase, cue condition comparisons within children with ASD and TD children (e.g., during baseline: ASD in arrow condition versus ASD in gaze condition, 3 comparisons per group, for 6 total).

Post hoc tests of the 3-way interaction exhibited no differences between children with ASD and TD children in their proportion of overall looking time during baseline, teaching, and test phases across cue conditions (*p*s > .21). Where significant differences did occur, was within groups for TD children. First, there was a cue condition effect during the test phase, where TD children spent a significantly smaller proportion of overall looking time during the gaze condition (PMM = .77, SE = .02) than the arrow condition (*PMM* = .87, *SE* = .02), *t*(456.85) = 3.89, *p* = .003, *PMM* difference = .10, *d* = -.70. It is unclear why TD children looked more overall at the video in the arrow condition during the test phase relative to the gaze condition, since the same test phase videos were used across cue conditions. No other comparisons between cue conditions in other video phases were significant within children with ASD (*p*s > .98) or within TD children (*p*s = 1.00).

A second significant effect was seen between video phases in the gaze condition, where TD children spent a significantly smaller proportion of overall looking time during test in comparison to baseline (*PMM* = .91, SE = .02), *t*(456.33) = .6.07, *p* < .001, *PMM* difference = .14, *d* = 1.1, and during test in comparison to teaching (*PMM* = .92, SE = .02), *t*(456.33) = 6.40, *p* < .001, *PMM* difference = .15, *d* = 1.26. The mean proportion of overall looking time of 74% during the test phase still seemed like children were looking during the majority of the test phase (raw *M* = .74, *SD* = .18, range = .28 - 1). In contrast, in the arrow condition, the proportion of looking time did not decrease across video phases for TD children, and the proportion of overall looking time was similar between baseline (*PMM* = .92, *SE* = .02), teaching (*PMM* = .90, *SE* = .02), and test phases (*p*s > .44). Children with ASD demonstrated a smaller proportion of overall looking time across video phases in both cue conditions, with less looking by the test phase. There were significant differences in children with ASD in the proportion of overall looking time between baseline (gaze *PMM* = .95, *SE* = .02; arrow *PMM* = .91, *SE* = .03) and test (gaze *PMM* = .79, *SE* = .03; arrow *PMM* = .77, *SE* = .03), and teaching (gaze *PMM* = .89, *SE* = .02; arrow *PMM* = .88, *SE* = .02) and test in both cue conditions (*p*s < .006). There were no significant differences between baseline and teaching phases in either group (*p*s > .33).

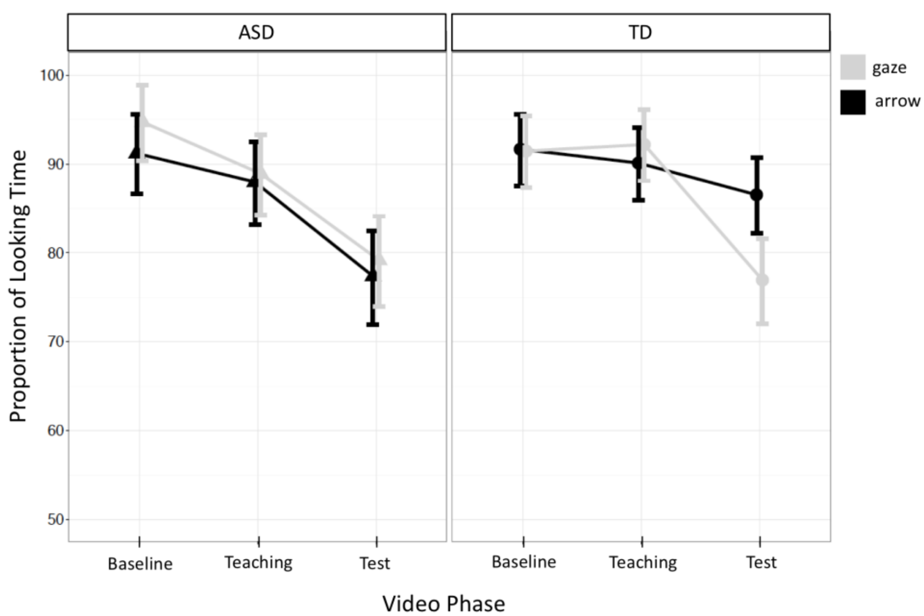
In summary, there were no group differences and as expected, both groups of children spent less time in their overall looking by the test phase, though TD children did not look less by the test phase in the arrow relative to the gaze condition. These results suggest any potential group differences in specific areas of interest (e.g., area of gaze, target), or cue condition differences within children with ASD, would not be due to any systematic differences in overall attention to video phases. However, within TD children, if there were any potential differences in specific areas of interest where during the test phase children looked less at areas of interest in the gaze relative to the arrow condition, this could be due to TD children spending less time looking overall at the gaze videos. No such differences were found in Chapter 2. Figure G 1 depicts the proportion of overall looking time across baseline, teaching, and test phases in each cue condition by group.

Table G 2. Action Prediction: Main Effects and Interactions of Overall Proportion of Looking time to the Scene

Factor	*df*	*F*	*p*
Cue condition (C)	1, 440.77	.01	.91
Group (G)	1, 40.96	.46	.50
Video phase (V)	2, 436. 14	53.84	<.001***
C x G	1, 440.77	4.92	.03*
C x V	2, 436.14	3.21	.04*
G x V	2, 436.14	2.33	.10
C x G x V	2, 436.14	2.96	.05^

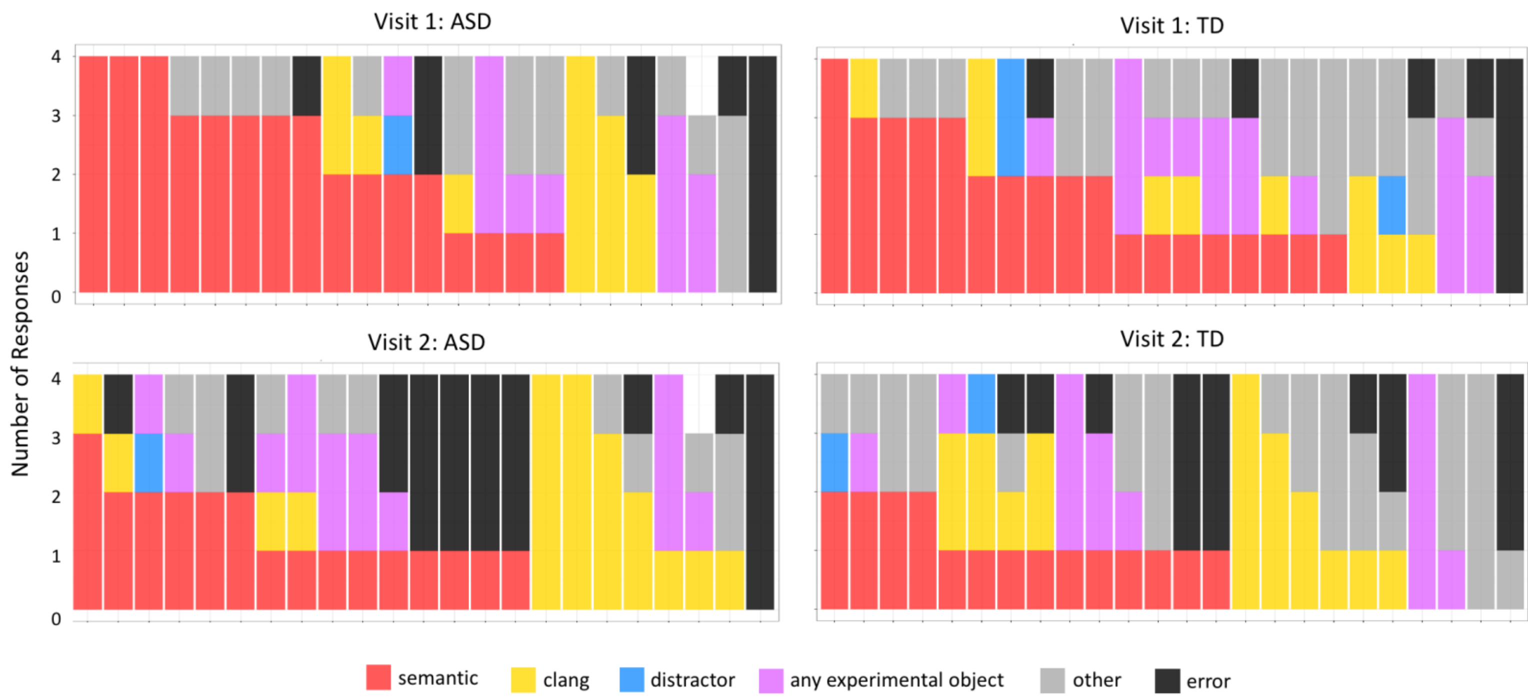
^ $p = .05$, * $p < .05$, *** $p <.001$

Figure G 1. Action Prediction: Predicted Marginal Means of the Overall Proportion of Looking time to the Scene Across Video Phases in Each Group of Children



This figure depicts that in both groups, children spent less time looking to video phases by the test phase in comparison to baseline and teaching phases, and the proportion of looking time does not differ between groups until the test phase. In the test phase, TD children spent more time looking during the arrow condition relative to the gaze condition. Points represent predicted marginal means and error bars are 95% confidence intervals.

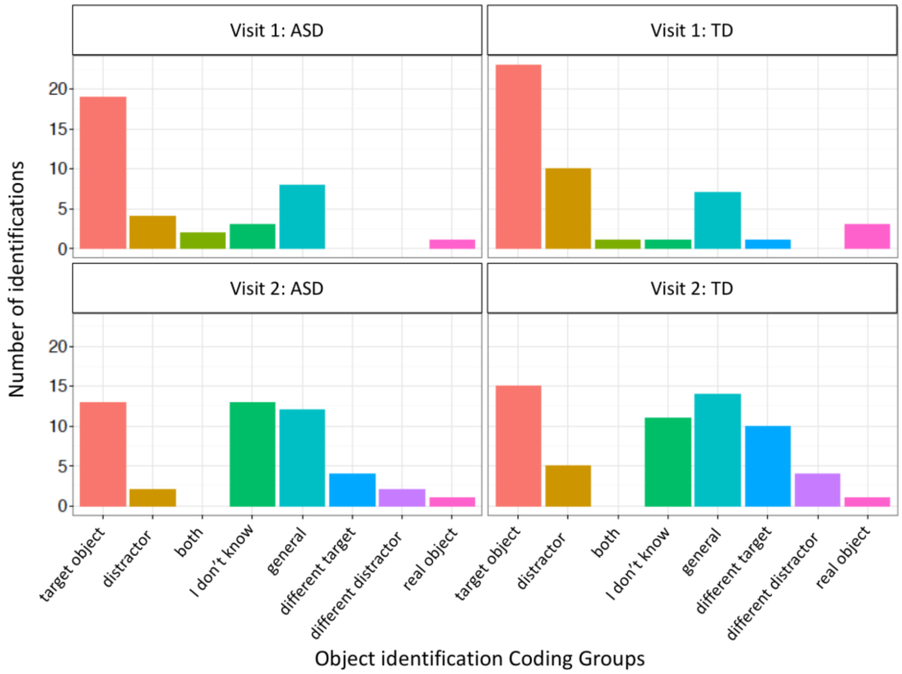
Appendix H. Word Association Responses for Individual Children



This figure depicts a count of the different types of responses coded for individual children. Each child provided four possible responses per visit (one word per target word), with a maximum of 8 responses per child. For presentation purposes, in each visit children's responses were organized from left to right for children who provided the most to the least semantic responses. Therefore, the location of one bar in one visit does not always correspond to the same child in the same location of the bar in the other visit. One ASD child has only three responses.

This figure depicts that both groups varied widely in the types of responses that were coded. When children did not give semantic responses, they gave a range of responses from coding groups of clang, distractor, any experimental object, other, and error, as can be seen by the different colors above. As seen by fewer red bars in visit 2 versus visit 1, semantic responses diminished from one week later in both groups of children. As semantic responses diminished in visit 2, more clang, other, and error responses can be seen (more colors corresponding to these coding groups in visit 2), reflecting difficulty remembering the target object one week later.

Appendix I. Object Identifications for Word Descriptions



This figure depicts a count of the different types of object identifications by the coder. The y axis represents the total number of object identifications summed across each group of children. Each child provided four possible responses per visit, with a maximum of 8 responses per child. The x axis represents the different types of descriptions by children, as determined by the coder.

Target object referred to descriptions that the coder determined as the correct target object. *Distractor* referred to descriptions regarding the distractor object that was paired with the target object. *Both* was used when the child appeared to include descriptions of both the target and the distractor object, thus the coder was unable to determine one object. *I don't know* referred to when the child simply said, "I don't know" and provided no other description. *General* was used when the child used only general terms (e.g., it's small), but there was no way to determine a specific object. *Different target* and *Different distractor* referred to when the child described another target or distractor object (e.g., a distractor associated with a different target). Finally, *real object* was used when the child described a real object (i.e., pagoune – binoculars, meaning that the child was describing the pagoune as if it were binoculars (e.g., *you use them to see*, rather than *you press one button and the other button pops up*).

At visit 1, most responses were of the target object. However, by visit 2, the number of responses were similar between coding groups of 1) the target object, 2) when children said they didn't know the object, or 3) general descriptions. In typically-developing children, there were also a similar number of responses when children were describing a different target object. Fewer identifications of the target objects by visit 2 reflects difficulty remembering the target one week later. Between groups, there were slight differences where in typically-developing children, there were more identifications in certain coding groups than for children with ASD (i.e., visit 1 distractor, visit 2 different target), but overall groups were similar in the distribution of the total number of identifications. These similarities between groups suggest that descriptions of objects by children with ASD were identified at similar rates across coding groups as descriptions of objects by typically-developing children.