Georgia State University
ScholarWorks @ Georgia State University

Psychology Dissertations

Department of Psychology

5-10-2017

Social Attribution in Toddlers At Risk for Autism Spectrum Disorder

Natasha Ludwig

Recommended Citation
Ludwig, Natasha, "Social Attribution in Toddlers At Risk for Autism Spectrum Disorder." Dissertation, Georgia State University, 2017.
http://scholarworks.gsu.edu/psych_diss/175

SOCIAL ATTRIBUTION IN TODDLERS AT RISK FOR AUTISM SPECTRUM DISORDER

by

NATASHA N. LUDWIG

Under the Direction of Diana Robins, PhD & Rebecca Williamson, PhD

ABSTRACT

Autism spectrum disorder (ASD) can now be reliably diagnosed in preverbal toddlers and early diagnosis is becoming more common since the development of early autism screening practices. However, the positive predictive value of widely used screening tools remains low, which leads to a high number of false positive cases requiring further evaluation. Given that access to specialists is limited, there is a pressing need to develop easily accessible and broadly applicable direct measures that will further streamline screening and diagnosis for at risk toddlers. The primary aim of the current study is to examine the utility of a novel, direct measure of social attribution in measuring this skill in toddlers with and without ASD and examining the utility of this measure in reliably identifying ASD in a sample of at risk toddlers with a broad range of verbal and cognitive abilities.

Participants include 35 toddlers considered at risk for an ASD (i.e., 15 with ASD, 20 with non-ASD delays; DD) and 22 typically developing (TD) toddlers. Children were presented with two versions of a nonverbal social attribution measure featuring a visual habituation-based violation of expectation paradigm; a live puppet show version previously studied in infant populations and a novel touchscreen adaptation. It was hypothesized that toddlers without a diagnosis of ASD would demonstrate evidence of social attribution whereas children with ASD would demonstrate reduced social attribution. Furthermore, it was predicted that performance would have clinical utility in predicting a diagnosis of ASD and symptom severity. Results indicated that no groups showed gross looking time differences evidencing social attribution, bringing into question whether this paradigm is appropriate for capturing social attribution in this age range. Despite this, toddlers in the TD group demonstrated evidence of social evaluation in the live puppet show task whereas toddlers within the ASD and DD groups did not. Differential habituation characteristics between the DD and TD groups suggest that other factors may have impeded success in the DD group. Future research is warranted to examine whether deficient social evaluation is specific to ASD or characterizes developmental delays more broadly. Findings have implications for future research examining theories of social attribution and informing the use of new technologies in toddler research and clinical tool development.

INDEX WORDS: Social attribution, social cognition, habituation, toddlers, autism, developmental delay

SOCIAL ATTRIBUTION IN TODDLERS AT RISK FOR AUTISM SPECTRUM DISORDER

by

NATASHA N. LUDWIG

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2017

SOCIAL ATTRIBUTION IN TODDLERS AT RISK FOR AUTISM SPECTRUM DISORDER

by

NATASHA N. LUDWIG

Committee Co-Chairs: Diana Robins & Rebecca Williamson

Committee: Lauren B. Adamson

Sarah Brosnan

Warren Jones

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2017

DEDICATION

To my brother Gianni. My little brother, my biggest inspiration.

ACKNOWLEDGEMENTS

I would like to acknowledge my dissertation committee for all of their support in the development and implementation of this project. I would also like to acknowledge the Developmental Neuropsychology Laboratory and the Developmental Laboratory at Georgia State University for helping to make this study possible. Specifically, I would like to thank Riane Ramsey, Shaina Aifuwa, Lashae Nichols, and Katie Suma for their help in data collection. I would also like to acknowledge all of the children and families who participated in this study and the funding sources including the Autism Speaks Weatherstone Predoctoral Fellowship, the Eunice Kennedy Shriver National Institute for Child Health and Human Development grants (R01HD039961, R01HD035612), and the Autism Speaks Targeted Research Award (#8368). Finally, I would like to acknowledge my family, friends, and mentor, Dr. Diana Robins, for the wealth of support they all have provided me throughout my years in graduate school.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASD = Autism spectrum disorder; DD = Developmentally delayed; PAT = Puppet Attribution Task; SAT = Social Attribution Task; TAT = Tablet Attribution Task; TD = Typically developing.

1 INTRODUCTION

Autism spectrum disorder (ASD) can be reliably diagnosed in preverbal toddlers, and early diagnosis is becoming more common since the development of early autism screening practices. However, the positive predictive value (PPV) of widely used screening tools remains low, which leads to a high number of false positive cases requiring further evaluation. Recent data suggest that 50-60% of children considered at risk for ASD based on the Modified Checklist for Autism in Toddlers, Revised (M-CHAT-R; Robins, Fein, & Barton, 2009) are diagnosed with an ASD upon further evaluation, and that most of the remaining children are diagnosed with another developmental delay (Robins et al., 2014). Given the expertise required for the differential diagnosis of ASD from other developmental delays in young children and the limited access to specialists due to long waitlists and dearth of specialists in certain regions of the US and the world, there is a pressing need to develop easily accessible and broadly applicable measures that will further streamline screening and diagnosis for at risk toddlers. Considering evidence that individuals with ASD demonstrate early deficits in social cognition, computerized measures of social cognitive functioning appropriate for toddlers with a broad range of cognitive and linguistic abilities may be a useful way to detect behavioral markers of ASD in toddlers and enhance ASD screening practices in pediatric primary care settings.

Social attribution is the social cognitive process by which typically developing individuals spontaneously attribute mental states to others based on movement cues alone (Heider & Simmel, 1944). Given that individuals across the autism spectrum demonstrate deficits in social attribution (Klin, 2000), and that performance on measures of social attribution is associated with the extent of social impairment in individuals with ASD (Boraston et al., 2006), disrupted social attribution may be a promising early behavioral marker of ASD. Therefore, the purpose of this study is to explore the utility of a novel, nonverbal, computerized measure of social attribution in predicting a diagnosis of ASD in a diverse sample of toddlers considered at risk for ASD based on the M-CHAT-R.

1.1 Theory of Mind (ToM) and mentalizing

Each day we are faced with a diverse array of social cues from our environment that we must perceive, integrate, interpret, and act upon. This complex cognitive processing is called social cognition (Carrington & Baily, 2009) and is a fundamental aspect of successful social interaction. Mentalizing is one aspect of social cognition involving the ability to infer mental states in others. The term theory of mind (ToM) is also used to describe this capacity and was first coined by Premack and Woodruff in their seminal paper (1978) discussing the sophistication of social cognition in non-human primates. Premack and Woodruff described ToM as the ability to utilize social cues to attribute mental states to others, including beliefs, feelings, and desires. In typically developing humans, mentalizing is an automatic and adaptive process as it facilitates the successful navigation of our social environment, whereas impairments in ToM can contribute to significant impairments in social interactions characteristic of psychological disorders such as ASD and schizophrenia (Baron-Cohen, 2000; Bora, Yucel, & Pantelis, 2009; Sprong et al., 2007; Yirmiya et al., 1998).

Since Premack and Woodruff's discussion of ToM in 1978, scientists and philosophers alike have considered the limits and bounds of this construct. Some of the first philosophers to discuss the modern concept of ToM (Bennett, 1978; Dennet, 1978) took a conservative stance, suggesting that ToM requires a representational understanding of the mind or the ability to appreciate a false-belief. Therefore, much of the subsequent research exploring ToM has focused on the ability to understand a false-belief (Bloom & German, 2000; Tager-Flusberg & Sullivan, 2000; Wimmer & Perner, 1983). In fact, some refer to traditional false-belief tasks as the “litmus test” for ToM (Frith & Frith, 1999). However, inherent in Premack and Woodruff’s initial definition is the notion that ToM also involves the attribution of mental states other than beliefs (e.g., desires, emotions). Given that the dichotomous nature of traditional false-belief tasks implies that an individual either has or does not have ToM, and that children generally do not pass these tasks until four years (Wimmer & Perner, 1983), a new generation of experimental measures have been developed to examine whether there are various psychological components that underlie mental state attribution, and whether there is variability in the emergence and expression of these capacities. These new ways to quantify ToM have contributed to the development of a number of theoretical accounts about the underlying processes implicated in ToM. Considering the scientific evidence supporting these various perspectives, it is likely that ToM is a complex construct that requires integration of many psychological and neurobiological processes. A recent review of the past 30 years of research in ToM (Call & Tomasello, 2008) supports that the modern understanding of ToM has certainly broadened from an all-or-nothing phenomenon to a dimensional understanding.

1.2 Social-perceptual ToM

Tager-Flusberg and Sullivan’s (2000) influential account of ToM proposes the need for a distinction between the traditional cognitive understanding of the mind as a representational system and the processes that allow us to make immediate mental state attributions based on perceptual information. They propose a componential theory, breaking down the mentalizing into *social-perceptual* and *social-cognitive* aspects. While the social-cognitive component encompasses the traditional interpretation of ToM, that is, the high-level cognitive processes necessary to understand a false-belief, the social-perceptual component includes the bottom-up processes that allow us to make more immediate mental state attributions about intentionality, goals, desires, personality, and emotions based on physical features of the stimuli such as facial expression, motion, and tone of voice.

From a developmental perspective, the social-perceptual component of ToM appears to emerge earlier than the social-cognitive component. Although findings from studies utilizing traditional false-belief tasks have shown that this capacity does not emerge until age four (Wimmer & Perner, 1983), social-perceptual ToM capacities are evident in the first two years of life. For example, children as young as three months look in the same direction as an adult's gaze (Hood, Willen, & Driver, 1998), and four-month-olds demonstrate distinct reactions to different facial expressions elicited by others during peek-a-boo play (Montague & Walker-Andrews, 2001). With regard to understanding intent, infants as young as five months infer that a person's reach is intentional (Woodward, 1998), and between nine and 12 months, babies understand others' intent to reference an object with a use of a pointing gesture (Woodward & Guajardo, 2002). By one year, babies utilize both negative emotional sounds and direction of gaze together to infer negative feelings about an object (Moses et al., 2001), and by fourteen months can utilize facial expression and gaze cues together to infer goals (Phillips, Wellman, & Spelke, 2002). Additionally, 12-month-olds infer intentions of inanimate objects that appear animate based on their interactive movements alone, suggesting the importance of movement cues in attributing mental states (Johnson, Slaughter, & Carey, 1998).

Neuroimaging evidence supports that the social-cognitive and social-perceptual components of ToM are distinct (Sax, Carey, & Kanwisher, 2004; Tager-Flusberg, 2000). For example, Gobbini and colleagues (2007) conducted an fMRI study to explore the neural involvement during a traditional false-belief task and a social-perceptual task, and discovered that the neural mechanisms involved in these ToM tasks are distinct. The false-belief task involved areas of the anterior paracingulate cortex, the posterior cingulate cortex/precuneus, and the temporo-parietal junction, which are brain areas often associated with false-belief tasks (Sax et al., 2004). However, the motion-based mentalizing task involved superior temporal sulcus, including a locus in the posterior superior temporal sulcus, as well as the frontal operculum and inferior parietal lobule, and the fusiform gyrus, which are areas commonly associated with facial perception and action understanding.

1.3 Motion

One physical feature important for social-perceptual mentalizing is human motion. Motion cues help us distinguish whether a moving object is animate or not (Tremoulet & Feldman, 2000) and can help us infer aspects of identity such as age, sex, and physiological state (Blake & Shifrar, 2007). Motion cues can also be socially relevant, that is, facilitate attribution of mental states such as goals, motivations, intentions, and emotions to others (Atkinson et al.; 2004; Schlottmann & Ray, 2009).

It is the specific kinematic elements of movement that contribute to social interpretation. The human visual system is an expert at detecting and interpreting social motion, even when human form information is limited as is evident from research using point-light displays (Blake & Shiffrar, 2007). In fact, when viewing full-body point-light displays, typically developing individuals automatically attribute emotion, motivations, and intentions to these stimuli.

Attention to social motion cues emerges very early in infancy (Fox & McDaniel 1982; Neri, Marrone, & Burr, 1998). As early as days after birth, typically developing children prefer to look at biological motion over non-biological motion, which supports the importance of experience with motion in early social development (Blake & Shiffrar, 2007; Simion, Regolin, & Bluf, 2008). In fact, it is thought that preference for biological motion is a developmental precursor to mental state attribution (Frith & Frith, 1999). Therefore, the use of tasks tapping the ability to use motion cues to make mental state attributions may have potential as a way to identify early deficits in social-perceptual mentalizing abilities.

1.4 Social attribution

Social attribution is a specific kind of motion-based mentalizing, occurring in the context of anthropomorphization, or the attribution of human features including mental states to geometric shapes based on their movements alone. This process has shown to be immediate and automatic, and is thus considered an aspect of social-perceptual mentalizing.

In a review by Scholl and Tremoulet (2000), the authors discuss motion characteristics that seem to facilitate mental state attribution in ambiguous visual displays. Scholl and Tremoulet paid homage to the original empirical study exploring this phenomenon conducted by Heider and Simmel in 1944 entitled, “An Experimental Study of Apparent Behavior.” Participants were presented with a silent movie featuring three geometric shapes scripted to follow social scenes (e.g., fighting, love triangle, chasing). Reflected in the title is the finding that the social intent of the shapes was apparent based on the shapes’ movements alone. In fact, typically developing individuals find it difficult to describe the movement of the shapes without using human terms, which suggests that the cognitive process implicated in our attempts to make sense of the specific movement stimuli is highly automatic. Interestingly, Heider and Simmel found that there was little variability in participant's verbal descriptions of what they saw: All but one participant (*n* = 34) described the shapes using social vs. physical terms, and participants attributed similar mental states to the figures including intentions, emotions, and personality traits. Of the 33 participants who described the scenes in social terms, 31 participants described the shapes as humans (i.e., they anthropomorphized the shapes), whereas the other two described them as birds. More recent studies have replicated Heider and Simmel's findings using adaptations of the original task in typically developing adults (Abell, Happe, & Frith, 2000; Klein et al., 2009; Klin, 2000). Additionally, research has shown that descriptions of social attribution scenes are similar across cultures (Rime et al., 1985), which suggests that motion cues are a universal aspect of social perception.

The impetus for Heider and Simmel's (1944) original study precipitated from their discouragement with the lack of literature at the time exploring physical characteristics that facilitate social interpretations of socially relevant stimuli. Based on their findings, Heider and Simmel explained that the anthropomorphic interpretations of the scenes were associated with the temporal contiguity and the spatial proximity of the visual stimuli. For example, animate objects were perceived as those that could move alone, whereas props did not. Additionally, when two shapes contacted each other, this could be characterized as a social interaction of some kind, but when an agent contacted a prop, it was typically interpreted that the shape controlled the prop (Klin, 2000). Subsequent work has been done to better understand the kinematic features of dynamic visual displays that mediate mental state attribution. Some have found that aspects of motion including temporal contingency with regard to changes in direction (Bassilli, 1976), the relationship between trajectories and velocities (Dittrich & Lea, 1994), the directness of movement towards each other, timing of closeness and the orientation of shapes towards each other (Gao, McCarthy, & Scholl, 2010; Gao, Newman, & Scholl, 2009; Gao & Scholl, 2011) all seem to elicit social interpretations. The Heider and Simmel stimuli are perceptually basic, yet humans quickly and efficiently make mental state attributions. Scholl and Tremoulet (2000) draw parallels between the way in which our visual system automatically works to recover gaps in perceptual stimuli to perceive physical properties and make sense of physical phenomena (e.g., perceiving a two-dimensional shape as three-dimensional based on shadowing), and the way in which we immediately make social interpretations about ambiguous stimuli based on motion cues in order to make sense of the scene.

1.5 The neural underpinnings of social attribution

Castelli and colleagues (2000) utilized PET to better understand functional activation associated with social attribution. Results showed that social attribution movies, similar to Heider and Simmel’s original scenes, selectively involve areas of the social brain (Frith, 2007) including the medial prefrontal cortex, temporoparietal junction (i.e., superior temporal sulcus), basal temporal regions (i.e., fusiform gyrus and temporal poles adjacent to the amygdala), and extrastriate cortex (i.e., occipital gyrus). Schultz and colleagues (2003) also explored the involvement of various brain regions in typically developing adults during an adapted version of Klin’s (2000) Social Attribution Task (SAT) using fMRI. Findings supported involvement of the amygdala, temporal pole, medial prefrontal cortex, inferolateral frontal cortex, superior temporal sulci, and the lateral portion of the fusiform gyrus, the fusiform face area. The involvement of the fusiform face area was especially interesting since the shapes in the SAT do not have faces, yet this area is known for robust activation during facial perception. Schultz and colleagues suggest that social interactions typically involve facial processing, and when social movement cues are apparent without human form cues, it is likely that this area is automatically recruited for social processing.

Heberlin and Adolphs (2006) described a neuropsychological case study exploring social attribution skills in an adult with bilateral amygdala damage. The patient described the shapes in asocial physical ways, despite typical visual perception and language skills. These findings further support the importance of the amygdala in social attribution. Consistently, interactive movement perception is associated with the amygdala and temporal poles (Santos et al., 2010). Interestingly, a study conducted in typically developing adults showed that anthropomorphization, involves similar underlying brain structures as attributing mental states to humans (Cullen et al., 2013). This suggests that individuals who have deficits in anthropomorphization in the context of social attribution tasks are likely to have deficits in real-world social interactions when attributing mental states to humans is also necessary. These findings support the use of social attribution measures to detect social impairment.

1.6 Early social attribution

Studies have shown that within the first year of life, babies are not only able to detect animate agents (Johnson, 2003) but are also able to infer goal-directed behavior (Woodward, 1998). However, although infants can identify the physical goal-related behavior, it does not necessarily mean that they attribute a mental state or disposition to the individual engaged in the behavior (i.e., that the individual *wants* to reach the goal). To explore whether babies have the capacity to attribute a mental state in the presence of goal-directed behavior, a series of studies were conducted using a modified social attribution task appropriate for nonverbal infants (Kuhlmeier, Wynn, & Bloom 2003). After watching brief computerized movies featuring geometric shapes helping or hindering another shape attempting to climb a hill, infants as young as 12 months make mental state attributions about the shapes' attitudes as evidenced by preference towards looking at novel scenes depicting a coherent continuation of the social interaction (i.e., climber floated to the top of the screen to be near the helper shape) compared to incoherent scenes (i.e., climber floated to the top of the screen to be near the hinderer; Kuhlmeier et al., 2003). These data suggest that infants utilize the mental state attributions (i.e., a disposition) they make about the shapes to make sense of the shapes' behavior. A follow-up experiment revealed that results could not be explained by a preference for positive associations between shapes (i.e., infants attributing a goal to the climber to get to the top of the hill and associating that when he reached the goal, the helper was near him, and when he didn't reach the goal, the hinderer was near him), but were in fact due to attribution of a disposition to the climber (Kuhlmeier et al., 2003). This was achieved by having the helper or hinderer move towards the climber during a novel test phase. Results from this follow up study showed that infants' looking time during these test trials did not differ and thus preference for the climber moving towards the helper indicated infants' attribution of a disposition to the climber.

These findings were replicated and extended to 10-month-old children in a study conducted by Hamlin, Wynn, and Bloom (2007) using a live puppet version of Kuhlmeier and colleagues' (2003) social attribution paradigm; 10-month-old infants evidenced surprise (i.e., longer looking time) when they witnessed the climber shape move towards the hindering shape instead of the helping shape. Notably, although Kuhlmeier and colleges (2003) and Hamlin and colleagues used a similar paradigm across studies it must be highlighted that infants actually looked longer to the expected event in Kuhlmeier study (i.e., the climber approaching the helper) and in contrast, looked longer towards the surprising event in the Hamlin study (i.e., the climber approaching the hinderer). Primary differences across these two studies include that the Kuhlmeier paradigm was computerized and used a test task featuring the climber shape floating in the air up toward the hinderer or helper, whereas the Hamlin paradigm was a puppet show with a more naturalistic test task featuring the climber shape moving towards the hinderer or helper positioned on the ground to the right or left. Nonetheless, there was a difference in looking time towards both events indicating that the infants engaged in social attribution.

Pilot data using a similar version of Hamlin and colleagues task (2007) in toddlers (*n* = 20, 20-29 months) revealed similar findings and lend support to the use of this paradigm in toddlers (Gonsiorowski, Hrabic, Ludwig, Williamson, & Robins, 2015). Fawcett and Liszkowski (2012) also replicated findings in 10-month-olds, and conducted a control condition where the shapes move in the same way up and down the hill, but never touch each other. Findings demonstrated that differences in looking time towards the expected vs. unexpected event could not be explained by low-level perceptual explanations. Additionally, Fawcett and Liszowski (2012) utilized eye-tracking techniques to examine whether babies predict whether a shape will move towards the helper or hindering shape, which has been shown to be a more complex measure than reacting based on a violation of expectation (Hood, Cole-Davies, & Dias, 2003). Babies directed significantly more anticipatory gaze toward the expected rather than the unexpected event, suggesting that they use the shapes' mental state to predict their behavior. Taken together, these studies support that children as young as 10 months anthropomorphize ambiguous shapes and make predictions about the shape's social behavior based on movement cues alone.

1.7 Social evaluation

Some studies have used nonverbal social attribution paradigms to explore a related aspect of social cognition involving young children's ability to make social evaluations about ambiguous visual stimuli based on motion cues. Whereas social attribution is the process of attributing mental states to others, social evaluation involves making judgments about another (e.g., if they are nice and helpful, or mean and harmful) based on observation of the other's behavior. Social evaluation is an adaptive process as it helps us determine who to affiliate with and who to stay away from. For example, a child who observes another child pushing other kids on the playground makes a social evaluation that the child is mean and chooses not to play with that child. Interestingly, infants make social evaluations about individuals based on observation of behavior even without active involvement in the social interaction they observe. For example, Premack and Premack (1997) found that infants make social evaluations about positive and negative interactions that they observe, and Hamlin and Wynn (2011) demonstrated that children as young as five months prefer to play with animal puppets that previously helped another animal puppet open a box containing a toy, or helped the puppet retrieve a dropped ball, over puppets who prevented the puppet from finding the toy or retrieving the ball.

Using the social attribution paradigm, Hamlin, Wynn, and Bloom (2007, 2010) explored whether babies take into account individuals' actions towards others in making their own social evaluations about the individual. This would be adaptive for early survival as it would help babies identify those who may be harmful vs. those who may be helpful to them. Hamlin and colleagues (2007) found that children six and 10 months old prefer to play with the helping shapes over the hindering shapes after viewing the social attribution puppet paradigm. Follow-up experiments suggest that this preference is based on both avoidance of hindering shapes and seeking out of helping shapes.

In order to explore whether this phenomenon extends downward to three-month-olds, the choice-task was adapted to measure preferential looking since three month olds are too young to reach for a shape to play with. Hamlin and colleagues (2010) found that older infants (six- and 10-month-olds) spent more time looking towards the shape they subsequently picked indicating that preferential looking time towards a particular shape is consistent with choosing that shape to play with. Consistent with findings in six- and 10-month-olds, three-month-old infants preferred the helping shape over the hindering shape and that preference was guided only by avoidance of the hindering shape, not seeking out of the helping shape, as was observed in the older infants. Given that three- and six-month-olds do not differentiate between mental state consistent and inconsistent behavior (Hamlin et al., 2007, 2010), the authors suggested that social evaluation develops earlier than the ability to infer a mental state in another. Despite this, three- and six-month-olds do anthropomorphize the geometric shapes based on movements alone, as evidenced by their social evaluations.

1.7.1 Habituation paradigms for toddlers.

Habituation is characterized as a progressive decreased response to presentations of a stimulus over time. Habituation is an adaptive learning response in an effort to maximize processing of unlearned vs. learned stimuli in the environment. This response is capitalized upon in infant habituation studies utilizing looking-time, as habituation is measured by decreased looking time towards a repeated stimulus. Visual habituation studies in infants have shown that habituation rates predict later outcomes such as intelligence (Bornstein & Sigman, 1986; Colombo, 1993; McCall & Carriger, 1993).

In the reviewed studies examining social attribution/evaluation in infant samples (i.e., Hamlin et al., 2007, 2010; Kuhlmeier et al., 2003) a habituation paradigm was utilized to present children with the social scenes. However, minimal habituation characteristics were reported (i.e., how many met habituation criterion within the maximum number of trials, mean number of trials to reach habituation criterion, mean looking times during habituation, number were excluded due to fussiness during habituation) so it is unclear how typically developing infants habituate to these particular scenes.

Despite the abundance of habituation studies in infants, there is a dearth of research utilizing this paradigm in toddlers. As previously mentioned, pilot data utilizing a violation of expectation puppet task similar to the task used by Hamlin and colleagues (2007, 2010) in a toddler population (*n* = 20, age range = 20-29 months) replicated findings based on the post-habituation measures of social attribution/evaluation (Gonsiorowski et al., 2015); however, habituation characteristics were not reported. This lends some support for use of this habituation-based paradigm in toddlers; however further research is required to understand its feasibility.

1.8 Social attribution in ASD

As reviewed, social attribution seems fundamental to human cognition, and weaknesses in this area may have cascading effects on social functioning. This has led theorists to propose that problems in this domain might contribute to the social deficits observed in individuals with ASD. In particular, the social communication deficits in ASD could be the result of a core deficit in social cognitive processing, which includes social attribution.

Many studies, by using traditional false-belief tasks, support the claim that individuals with ASD demonstrate specific deficits in mentalizing compared to typically developing children and children with other delays (Baron-Cohen, Leslie, & Frith, 1985); however other studies have shown that some individuals with ASD, especially those who are higher functioning, succeed in these tasks despite impairment in real-world social functioning (Bauminger & Kasari, 1999; Bowler, 1992; Peterson, Slaughter, & Paynter, 2007, Senju et al., 2009; Shamay-Tsoory, 2008). Research suggests that those with high functioning ASD who succeed on false-belief tasks may be using compensatory strategies, which are not available or useful in real-world social situations. For example, individuals with Asperger's disorder did not use mental state terms when asked to explain how they successfully solved a false-belief task (Bowler, 1992), and they tend to use different parts of the brain compared to typically developing individuals, despite accuracy, suggesting alternative strategy use (Schultz et al., 2003). Alternate strategy use is also supported by evidence that verbal skills are a strong predictor of performance on these tasks, suggesting that individuals use verbal scaffolding to succeed (Tager-Flusberg &

Sullivan, 1994; Yirmiya et al., 1998). Additionally, traditional false-belief tasks are typically presented in an explicit problem-solving format, which is not representative of real-world social situations where individuals must spontaneously attend to, identify, interpret, and utilize ambiguous social information.

Also important is the dichotomous nature of ToM tasks, which likely contributes to the low test-retest reliability of these measures (Klin, 2000) and implies that ToM is an all-or-none phenomenon, which is likely not the case (Yirmiya et al., 1998). Taken together, these data suggest that traditional false-belief tasks may not capture the full range of ToM, and may contribute to the inconsistent findings across the autism spectrum.

Although individuals with mild forms of ASD may succeed in solving social-cognitive tasks such as measures of false-belief, they tend to show deficits during social-perceptual ToM tasks, which is likely due to reduced opportunities for use of more cognitive compensatory mechanisms (Baron-Cohen, Joliffe, Mortimore, and Robertson, 1997; Klin, 2000; Klin, Baron-Cohen et al., 1997; Tager-Flusberg and Sullivan, 2000). Additionally, because social-perceptual mentalizing skills emerge earlier than the social-cognitive components, social-perceptual tasks have potential to detect deficits earlier. Taken together, social-perceptual mentalizing tasks may be more valid measures of ToM in individuals across the autism spectrum throughout the lifespan.

In response to the limitations of traditional ToM tasks, the Social Attribution Task (SAT), a social-perceptual ToM task requiring utilization of motion cues for mental state attribution, was developed as a valid measure of ToM (Klin, 2000). The SAT is an adaptation of Heider and Simmel's (1944) scenes featuring ambiguous geometric shapes moving in social ways. Klin's adaptation of the traditional Heider and Simmel (1944) stimuli included minor adaptations to make the stimuli appropriate for individuals with developmental disabilities as well as the development of a coding scheme that could capture the variability in social attribution responses. The task reduces the possibility of verbal scaffolding by limiting the explicit nature of the instructions and task demands because participants are required to spontaneously search for social meaning in the ambiguous displays based on motion cues alone. Typically developing individuals automatically anthropomorphize the shapes and attribute mental states to them as is evident by their verbal descriptions of the scenes (Klin, 2000).

Klin also recruited a sample of adolescents and adult participants with autistic disorder and Asperger's disorder who had previously passed a traditional false-belief task, despite real-world social impairments. Individuals in both ASD groups described the movies in significantly less social ways compared to a typically developing sample. Specifically, individuals with ASD described fewer of the social elements of the scenes, provided more information unrelated to the social plot, attributed fewer personality traits to the shapes, and did not use mental state attribution terms as frequently in their descriptions compared to controls. Importantly, these deficits were not associated with verbal IQ or metalinguistic skills.

A follow-up study compared performance on the SAT with a novel Physical Attribution Task (PAT) in a group of individuals with ASD and a typically developing control sample (Klin & Jones, 2006). Results supported reduced social attribution in ASD compared to intact physical attribution skills. This finding has been replicated in a number of behavioral studies exploring performance on similar adaptations of the original Heider and Simmel task (1944) in ASD populations (Abell, Happe, & Frith, 2000; Bal et al., 2013; Boraston et al., 2006). Additionally, evidence suggests that performance on forced-choice social attribution measures correlates with impairments in social interaction as measured by the Autism Diagnostic Observation Schedule (Boraston et al., 2006). Findings also extend to young children with autism. For example, four to six- year-olds with autism take longer to learn to discriminate animate from inanimate geometric shapes based on movements alone, compared to children with developmental delays and typically developing children (Rutherford, Pennington, & Rogers, 2006), and children with autism as young as five are able to detect differences between mechanical vs. intentional movements of the shapes, but they make less reference to coordinated actions between shapes in their verbal descriptions, compared to typically developing children (Bowler & Thommen, 2000).

1.9 Social evaluation in ASD

Research suggests that individuals with ASD have difficulty making and sustaining friendships, which is likely influenced by general social cognitive deficits (Baumenger & Shulman, 2003). Despite the claim that the ability to make social evaluations about others is integral to having successful social relationships, there is minimal research exploring the way in which children with ASD make social evaluations about others. Findings from one pilot study (Gonsiorowski et al., 2015) utilizing the traditional social attribution habituation paradigm in a sample of toddlers showed that toddlers with ASD demonstrated reduced social evaluation (i.e., chose a helping and hindering shape at chance) compared to typically developing toddlers (i.e., 79% chose the helper), despite preserved evidence of social attribution in both groups as measured by surprise when the climber shape moved towards the hindering shape.

1.10 Neural evidence of disrupted social attribution in ASD

Evidence from neuroimaging studies also support reduced social attribution in individuals with ASD. Several neuroimaging studies have revealed that individuals with ASD demonstrate unique patterns of activation during social attribution tasks compared to typical controls; whereas typical controls demonstrate selective activation in brain areas associated with mentalizing (including the medial prefrontal cortex, superior temporal region, fusiform gyrus and temporal poles) while engaged in tasks of social attribution, individuals with ASD show reduced activations in these areas of the social brain (Castelli et al., 2000; Castelli et al., 2002; Ludwig et al., in prep; Schultz, et al., 2003).

1.11 A note about the social cognition and social motivation hypotheses of ASD

Although deficits in mentalizing are often observed in ASD, there is some controversy in the field about whether the social impairments in ASD are best explained by a core deficit in social cognition (i.e., ToM) or a core deficit in attention to social information, known as the social motivation hypothesis (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012). Notably, the fundamental difference between these theories relates to causality of functional deficits. Specifically, the social cognition hypothesis posits that individuals with ASD are interested in social information, but due to core difficulties in social processing, eventually lose interest. In contrast, the social motivation theory posits that individuals with ASD demonstrate a core deficit in social attention given a reduced neurobiological sensitivity to social reward, which contributes to downstream difficulties in social cognition. For example, research suggests that very young children with ASD show reduced fixation on people and faces during dynamic clips and reduced preference to socially salient sounds (Klin, 1991; Klin et al., 2002). Furthermore, very young children with ASD also show a preference for nonsocial physically contingent movement vs. socially relevant biological motion (Klin et al., 2009) and dynamic geometric images vs. social images (Pierce et al., 2011). Chevallier and colleagues (2012) suggest that deficits in social motivation precede deficits in social cognition; however, the authors also highlight that ToM has not adequately been studied in infants who later develop ASD. Although the current study does not aim to support or refute either of these hypotheses, it does aim to examine the utility of a measure of social attribution in toddlers, which may be appropriate for use in infant studies of mentalizing in the future.

1.12 Habituation paradigms in ASD, non-ASD delays, and typical development

Although there are few studies examining visual habituation characteristics in typically developing toddlers, there have been some studies examining visual habituation in toddlers with and without delays. Specifically, a recent study by Vivanti and colleague's (2017) examined toddlers' habituation to novel vs. repeated non-social stimuli (i.e., geometric shapes) presented simultaneously on a screen. Notably, toddlers with ASD required more trials to habituate to the repetitive stimuli compared to typically developing and toddlers with Williams' syndrome; however, the percentage of toddlers who did not reach habituation during the maximum number of presentations were not reported. Furthermore, toddlers with ASD also demonstrated reduced looking towards both the novel and repetitive stimuli compared to typically developing toddlers and those with William's Syndrome suggesting a decreased attention to novelty in this paradigm. Another study examined habituation to social stimuli in toddlers with and without ASD utilizing a paradigm well studied in infants (Webb et al., 2010). Specifically, toddlers were habituated to visual stimuli of a faces or houses and there were no maximum number of trials to reach habituation. Results indicated that toddlers with ASD with high symptom severity took longer to habituate to social vs. non-social stimuli and took longer to habituate to social stimuli compared to toddlers with ASD with low severity of symptoms, toddlers with non-ASD delays, and typically developing toddlers. Notably, this toddler study also presented novel and habituated stimuli to toddlers after habituation and found no novelty preference (i.e., increased looking time towards the new stimuli) in the typically developing group or the ASD group. This lack of novelty preference in the TD group suggests that the habituation paradigm and/or the novelty outcome procedure does not work the same way in toddlers as it does for infants. Taken together, it appears that toddlers with ASD may demonstrate differential habituation compared to typically developing toddlers and toddlers with non-ASD delays, but it is questionable whether visual habituation paradigms in toddlers with and without delays function the same way as they do in infant studies.

1.13 Touchscreen technology as an experimental paradigm in the study of ASD

Given that access to clinicians specializing in early diagnosis of ASD is limited, there is a pressing need to develop easily accessible and valid measures that can be administered in a variety of contexts and with children with a broad range of abilities in order to further streamline screening and diagnosis of ASD in toddlers. Considering this, the development of computerized methods that capture behavior in toddlers may be a useful technology in meeting this need.

Considering the prevalence of motor dexterity issues in ASD, especially very young children, touchscreen technology has become a popular learning tool in this population (Vasavi & Susmitha, 2012). For example, touchscreen tablets have been used in developing portable augmented communication systems for use in children with autism and have shown successful in helping children with autism build language (Samnath et al., 2012). However, given that touchscreen computers are relatively new, research on the use of touchscreen technology in experimental paradigms conducted in young children with disabilities is limited despite research demonstrating the feasibility of touchscreen devices in experimental and clinical settings (Stephenson & Limbrick, 2015). A recent review of literature examining use of touchscreen in individuals with disabilities showed that most studies were related to use of touchscreens for speech generation, prompting, and leisure activities and stresses the lack of literature examining touchscreen use in young samples (Stephenson & Limbrick, 2015).

A pilot study conducted structured interviews with doctors and therapists working with children with ASD and results indicated that 85% of experts believed that youth with ASD prefer to play with touchscreen devices over regular computers (Rias & Dehkordi, 2014). Additionally, there has been one study examining the use of a touchscreen computer in examining whether children with autism can detect animacy in the movements of geometric shapes (Rutherford, Pennington, & Rogers, 2006). The study included four- and five-year-old children with autism, developmental disabilities, and typically developing children and asked the children to touch the one of two geometric shapes moving on a screen that was animate, and a control task, which asked the children to touch the one of two heavier objects. Children with autism showed deficits in the ability to detect animacy compared to the other two groups, despite adequate performance on the weight task. One recent study examined the utility of a range of assessment methods, including a touchscreen task to determine receptive language skills in a group of minimally verbal children and adolescents with ASD (Plesa Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2016). Specifically, assessments of receptive language included traditional standardized direct assessment methods, parent report methods, and two experimental assessment tasks, one using eye tracking and another using a touch screen. Notably a number of children were able to demonstrate receptive language skills on the touchscreen task even though they were unable to demonstrate knowledge on the direct standardized assessment, which included many of the same words. This demonstrates that touchscreen methods may be a useful way to assess individuals with a broad language of language abilities who may not otherwise be able to be assessed using traditional methods. With regard to feasibility, the eye-tracking task was associated with greater amount of unreliable data compared to the touchscreen task. Taken together, these data suggest that touchscreen paradigms may be a useful way to study individuals with ASD and may have potential for clinical application.

With regard to the broader screen media literature, children with ASD spend approximately 62% more time playing with screen activities than all non-screen activities, and children with ASD play video games significantly more hours than their typically developing siblings (Mazurek & Wenstrup, 2012). Another questionnaire study suggested that children with autism spend more time engaged with electronic screen media over any other leisure activity (Shane et al., 2008). Additionally, these children also preferred animated screen media over other kinds of screen media. This fits with evidence from typically developing children suggesting that perceptually salient aspects of learning environments including sound effects and action increase attention and processing (Calvert, 1999). Studies comparing use of teacher or computer word acquisition interventions in children nine and under (Heimann, 1995; Moore

& Clavert, 2000) found that children with ASD learned more words from the computer instruction over teacher instruction and children with ASD paid more attention to and were more motivated by the computer intervention. Additionally, one study shows that while individuals with ASD demonstrate deficits on traditional cognitive tasks, these deficits are not shown on computerized versions of the same task (Landry & Al-Taie, 2016). Taken together these findings suggest that screen media may be highly engaging and may even scaffold learning or performance in individuals with ASD.

1.14 Present study

Given the ability to measure early emerging mental state attribution and social evaluation using nonverbal versions of the social attribution paradigm in preverbal infants, combined with evidence that social-perceptual aspects of social cognition are specifically disrupted early in development across the autism spectrum, deficits in this fundamental aspect of social processing may be a promising early behavioral marker of the disorder. Additionally, given that these abilities come online very early in development (mental age of 10 months for social attribution and three months for social evaluation; Hamlin et al., 2007, 2010), the social attribution paradigm is appropriate for children with a range of verbal and cognitive abilities, making this kind of task ideal for broad clinical applicability. Therefore, this study aims to evaluate the utility of a nonverbal social attribution task in differentiating toddlers with ASD from toddlers with non-ASD delays and typically developing children based on performance on social attribution and social evaluation measures. Furthermore, comparing performance and level of task engagement across live (a Puppet Attribution Task; PAT) and computerized (a Touchscreen Attribution Task; TAT) modalities may inform the feasibility of touchscreen technology in clinical tool development and in the study of child development more broadly. This study also aims to explore habituation specific characteristics that may inform the utility of habituation paradigms such as the PAT and TAT in toddler populations and further build upon the very small body of literature examining potential differences in habituation characteristics that may differentiate toddlers with ASD from toddlers with non-ASD delays and typical development.

Although implementation of this paradigm may not be feasible as a clinical tool to facilitate screening given the reliance on gaze monitoring, which requires expertise, this study seeks to provide the basic science necessary to inform the development of more implementable designs using social attribution measures and touchscreen technology. Additionally, this study will also provide data about use of a new way to explore social cognition in preverbal infants/toddlers and individuals with ASD through the utilization of new touchscreen interfaces.

1.15 Specific aims

1.15.1 Aim 1

Given the dearth of research utilizing habituation paradigms in toddler populations, it is important to examine habituation characteristics during the learning phase of the task. From a methodological perspective, it is important to determine whether the social attribution paradigm used in previous infant studies (Hamlin et al., 2007, 2010; Kuhlmeier et al., 2003) is feasible for use in toddlers. Specifically, it is important to know whether this paradigm is successful in habituating typically developing toddlers as well as toddlers with ASD and non-ASD delays to the stimuli as this will inform use of habituation paradigms in future toddler research as well as clinical tool development. With regard to the current study, examining habituation characteristics is necessary in informing the integrity of findings. Specifically, if children did not habituate to the scenes, this puts into question whether subsequent performance on measures of social attribution are valid. Furthermore, examining habituation characteristics by diagnosis will inform whether group differences in performance on measures of social attribution are accounted for by differential habituation rather than differences in social attribution. Beyond this, examination of habituation characteristics may also inform the broader body of research examining habituation differences in children with ASD and non-ASD delays compared to typically developing children and whether these differences may be a useful behavioral marker in early identification.

Aim 1 explores whether toddlers habituated (i.e., met habituation looking-time criterion within 12 habituation trials) to the stimuli presented in a Puppet Attribution Task (PAT) and Touchscreen Attribution Task (TAT) and whether habituation characteristics differed by diagnostic group. This is accomplished by examining whether the number of toddlers who reached habituation criterion differed based on task modality and diagnosis, whether the number of exposures required to meet habituation criterion differed by task modality and diagnosis, and whether there were differences in engagement between tasks and diagnostic groups. It was hypothesized that a similar proportion of toddlers in each group would reach habituation criterion and that this would not differ by modality, but that children with ASD would require increased exposures to reach habituation criterion compared to the TD and DD groups. It was also predicted that children with ASD would demonstrate reduced engagement (i.e., less average looking time per habituation trial) compared to the TD and DD groups for the

PAT, but would demonstrate increased engagement on the TAT compared to the TD and DD groups.

1.15.2 Aim 2

Aim 2 examines performance on measures of social attribution and social evaluation during the PAT. Specifically, Aim 2 serves to determine whether performance on the PAT would differentiate ASD from non-ASD delays, in a sample of at-risk toddlers, and typical development. It was hypothesized that children without ASD (i.e., DD and TD) would demonstrate evidence of social attribution, as demonstrated by increased looking towards an unexpected event compared to an expected event (i.e., PAT Looking Time Test), and evidence of social evaluation, as demonstrated by preference for choosing a helping over hindering character (i.e., PAT Choice Test), whereas children with ASD would demonstrate reduced evidence of social attribution and social evaluation on these measures. Additionally, it was predicted that performance on these measures would predict a diagnosis of ASD and ASD symptom severity.

1.15.3 Aim 3

Aim 3 examines performance on measures of social attribution and social evaluation during the TAT. Specifically, Aim 3 serves to determine whether performance on the TAT would differentiate ASD from non-ASD delays, in a sample of at-risk toddlers, and typical development. It was hypothesized that children without ASD (i.e., DD and TD) would demonstrate evidence of social attribution, as demonstrated by increased looking towards an unexpected event compared to an expected event (i.e., TAT Looking Time Test), and evidence of social evaluation, as demonstrated by preference for choosing a helping over hindering character (i.e., TAT Choice Test), whereas children with ASD would demonstrate reduced evidence of social attribution and social evaluation on these measures. Additionally, it was predicted that performance on these measures would predict a diagnosis of ASD and ASD symptom severity.

1.15.4 Aim 4

Little is known about the use of touchscreen devices in experimental paradigms for toddlers with and without delays and research is needed to inform utility of this relatively new technology for future research and clinical tool development. While Aim 1 addresses potential differences in habituation characteristics across modality, Aim 4 serves to examine potential differences in performance on measures of social attribution across task modality. This is important not only to learn more about the utility of the paradigm, but also to understand whether modality-based differences may impact performance on measures of social attribution and social evaluation for all toddlers or differentially by diagnostic group. For example, should individuals with ASD demonstrate increased engagement with the TAT compared to the PAT during habituation trials as predicted, it is important to learn whether this is associated with improved performance on the TAT. In this example, increased engagement may or may not be the reason for improved performance, but this finding would present additional research questions warranting follow-up. Therefore, this exploratory aim examines whether differences in performance on measures of social attribution (i.e., Looking Time Test) and social evaluation (i.e., Choice Test) emerged on the PAT and TAT and whether this differed based on diagnostic group. It was hypothesized that modality of task presentation would not be associated with changes in performance on social attribution and social evaluation measures across diagnostic groups.

2 METHODS

2.1 Participants

2.1.1 Recruitment

All toddlers in the ASD and DD groups were recruited through ongoing studies of the Modified Checklist for Autism in Toddlers, Revised plus Follow-Up (M-CHAT-R/F; Robins, Fein, & Barton, 2009) at Georgia State University (GSU), funded by NICHD and Autism Speaks. Recruitment of toddlers in the TD sample was accomplished through the M-CHAT-R/F studies, another study of longitudinal development at GSU funded by NICHD, or word of mouth. All toddlers included in the TD sample received the M-CHAT-R/F either through participation in another study at GSU or through participation in the current study and did not show risk for ASD on this measure.

2.1.2 Inclusionary/exclusionary criteria

All toddlers recruited in this study were between the age of 15 and 36 months English speaking, and parent/guardian consent for participation was obtained. Toddlers with severely impaired sensory or motor functioning that would interfere with experimental procedures were excluded.

For the ASD and DD groups, no other comorbidities were excluded in order to maximize generalizability of findings to the broad range of children with developmental delays. All toddlers included in the ASD or DD groups were considered at risk for ASD based on screen positive status, doctor concern, or on the Screening tool for Autism in Toddlers and Young Children (STAT; Stone et al, 2000; a second screening tool that was used in the larger screening study to detect ASD cases missed by M-CHAT-R/F), and received a comprehensive diagnostic evaluation at GSU through the M-CHAT-R/F studies. Toddlers were excluded from the sample if they did not receive a DSM-5 diagnosis after evaluation based on a positive M-CHAT-R/F screen.

For the TD group, history of developmental delays, neurological problems, or psychological diagnoses was exclusionary. All toddlers in the TD group did not show risk for ASD on the M-CHAT-R/F.

2.1.3 Sample

This study recruited 116 toddlers from February 2015 through March 2017. The experimental measures of social attribution (i.e., PAT and TAT) were piloted on the first 12 participants. Because the task design was changed after the pilot phase, data collected from the first 12 participants were excluded from analyses. Of the 104 remaining participants, 14 were considered at risk for autism based on the M-CHAT-R/F and were referred for comprehensive diagnostic evaluation through the M-CHAT-R/F studies but did not receive a DSM-5 diagnosis upon evaluation. They were excluded from analyses. Four additional children were erroneously recruited despite meeting exclusionary criteria (i.e., deafness (n = 1), outside age range (n = 1), high ADOS-2 score despite GDD diagnosis (n = 1), study staff concerned about development for a child in the TD group (n = 1) and were also excluded. An experimental error occurred during nine administrations (e.g., computer malfunction, sound malfunction, administration error) and data collected from these children were excluded.

Of the 77 remaining toddlers, one was consented, but did not participate in any sessions as this child was lost to follow up, and 16 toddlers participated in one session, but were lost to follow up. Descriptive characteristics of these children are reported in Table 1. Three children did not tolerate the task during Habituation Phase and were considered “fuss outs.” One child in the TD group fussed out during the PAT (first session then lost to follow up) and two children during the TAT (one child in the TD group during the first session and was lost to follow up, one child in the DD group during the second session and PAT was completed). Only toddlers who successfully completed (i.e., defined as full habituation data collected) both experimental measures of social attribution (i.e., PAT and TAT) were included in the final sample.

Table 1. *Characteristics of Toddlers Lost to Follow Up*

Variable	Total (*n* = 16)	ASD (*n* = 2)	DD (*n* = 6)	TD (*n* = 8)
Age at Session 1 (months)				
Mean	21.87	20.08	22.01	22.2
Range	18-28	20	19-28	18-27
Standard Deviation	3.13	0.44	3.57	3.29
Sex				
Male (%)	13 (76.5)	2 (100.0)	4 (66.7)	7 (87.5)
Female (%)	3 (17.6)	0 (0.0)	2 (33.3)	1 (12.5)

Fifty-seven toddlers (n_{male}=34) between the ages of 15 and 29 months of age at the time of Session 1 (*m*=21.61, *SD*=2.96) completed the PAT and the TAT and were included in the final sample (i.e., n_{ASD}=15, n_{DD}=20, n_{TD}=22). All toddlers in the ASD and DD groups screened positive on the M-CHAT-R/F, with the exception of two children in the DD group; one of whom was considered at risk due to doctor concern and the other who screened positive on the STAT. All children within the TD group received M-CHAT-R/F scores in the screen negative range. Fourteen toddlers in the DD sample had a diagnosis of GDD (70%) and the remaining six were diagnosed with LD (30%). Nine toddlers in the TD sample were recruited through the M-CHAT-R studies (40.9%), 12 through another study of longitudinal development at GSU (54.5%), and one through word of mouth (4.5%).

2.2 Experimental procedures

All toddlers in the ASD or DD groups were invited for a comprehensive diagnostic evaluation through the M-CHAT-R/F (Robins, Fein, & Barton, 2009) studies at GSU. The comprehensive evaluation included a caregiver reported history, cognitive/developmental testing (i.e., Mullen Scales of Early Learning; MSEL; Mullen, 1995), assessment of adaptive functioning (i.e., Vineland Adaptive Behavior Scales, Second Edition; Sparrow, Cicchetti, & Balla, 2005), and ASD specific diagnostic assessments based on direct observation (i.e., Autism Diagnostic Observation Schedule, Second Edition; ADOS-2; Lord et al., 2012), and parent report of behavior (i.e., Toddler ASD Symptom Interview; TASI; Barton et al., 2012). Final diagnoses were made based on clinical judgment based on DSM-5 criteria. For children in the ASD and DD groups, clinical data from participation in the M-CHAT-R/F studies were shared to categorize toddlers into diagnostic group. Specifically, toddlers diagnosed with ASD were included in the ASD group and toddlers diagnosed with Global Developmental Delay (GDD) or Language Delay (LD) were included in the DD group. Furthermore, MSEL data was used to characterize the sample and M-CHAT-R/F and ADOS-2 scores were used to explore whether performance on these measures relates to performance on experimental measures of social attribution in the current study.

Participation in the current study occurred during two sessions on separate days. Mean number of days between Session 1 and 2 was 14.04 (*SD* = 14.60; median = 7.00). Notably, time between sessions was greater than one month for only four participants (n_{ASD} = 1, n_{TD} = 3). Session 1 involved consent, administration of the PAT or TAT (task order counterbalanced across sessions), and administration of the Screen Media Questionnaire (SMQ; as well as additional questionnaires as part of other studies) to parents. Session 2 involved administration of the PAT or TAT (whichever was not administered during Session 1), de-briefing, and compensation for full study completion (i.e., $20). Toddlers in the TD group who did not receive the M-CHAT-R/F (i.e., toddlers recruited through word of mouth) or the MSEL (i.e., toddlers recruited through word of mouth or though M-CHAT-R/F studies) through other studies at GSU were also administered the M-CHAT-R/FU and/or the MSEL Visual Reception scale as part of study participation during Session 1.

Most toddlers were recruited after their participation in another study at GSU. If the family agreed to participate, Session 1 of the current study was run that day. Toddlers in the ASD and DD groups typically returned to GSU to complete their participation in the other study, and Session 2 of the current study was typically added to that session. A smaller group of toddlers in the ASD group and TD group returned to GSU to complete Session 2 of the current study only. Toddlers in the TD group who were recruited through another study at GSU typically returned to GSU to complete Session 2 of the current study only. For toddlers in the TD recruited through word of mouth (*n* = 1), both sessions were scheduled independently of participation in any other study at GSU.

2.3 Experimental measures of social attribution

2.3.1 Puppet Attribution Task (PAT)

The PAT was modeled based on previous studies examining social evaluation in typically developing children (Hamlin et al., 2007). The task includes a Habituation Phase as well as two test phases, which occurred immediately after habituation: the Looking Time Test Phase and the Choice Test Phase. The Looking Time Test always preceded the Choice Test. Performance on both test phases are intended to demonstrate whether children understood the social context of the scenes presented during the Habituation Phase and are considered primary measures of social attribution in the current study.

2.3.1.1 PAT Habituation Phase

Toddlers were habituated to helping and hindering social puppet scenes featuring a green hill and three colored shape puppets (circle, square, and triangle) displayed on a puppet stage. See Figure 1 for a static picture of a PAT Habituation scene. Notably, research has shown that typically developing children less than 12 months old require googly eyes on the shapes to demonstrate success in similar tasks (Hamlin et al., 2007; Kuhlmeier et al., 2003); however, because the presence of eyes provides increased scaffolding for mental state attribution, there were no eyes on the shapes in order to study the purest form of social attribution for the purposes of this study.

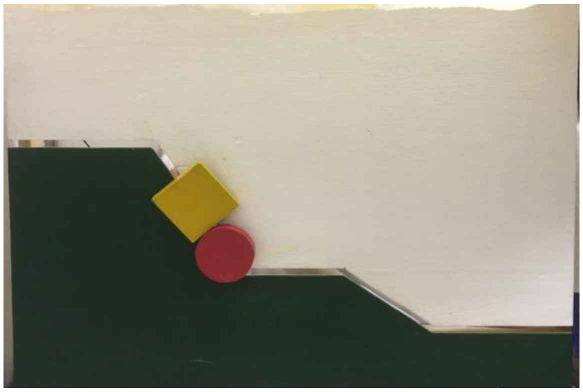

Figure 1. Static picture of a PAT Habituation Phase scene.

The scenes were shown on a puppet stage made of wood (33" x 24.5") containing a green incline. The shape puppets were made of wood and were all 3.5 square inches in size. Each scene featured a red circle as the climber shape and a yellow and blue triangle and square as the helper and hinderer (colors and shape type counterbalanced with constraint). Toddlers sat either on their parent/guardian’s lap or in a highchair with their parent standing next to them (based on parent preference). Parents were instructed not to interfere with their child’s behavior during the task, but could say “Wow!” or “Cool!” to express interest with the experimental environment. Parents were explicitly instructed not to instruct their child to “Look!” at anything during the task. Toddlers were seated 3.5 feet away from the puppet stage.

Each Habituation trial consisted of a scene presentation phase followed by a looking time phase. An experimenter, who was out of sight, controlled the climber, helper, and hinderer shapes. At the start of each habituation trial, the curtain was removed to reveal the puppet stage with the climber shape at the bottom of the hill, which began the scene presentation phase. In all scene presentations, the climber shape attempted to climb up the hill, but fell back down twice. In the helping trials, the helper shape came on the stage after the climber fell twice and pushed the climber up the hill and then the climber wiggled (i.e., in “celebration”). In the hindering trials, the hindering shape came on the scene after the climber fell twice and pushed the climber down the hill during the climber’s third attempt up the hill. The number of touches, the speed of touches, and the duration of touches between the helper or hinderer and the climber were the same to minimize differences across conditions. Each scene presentation lasted approximately 10 seconds.

A bell rang to end the scene presentation phase and begin the looking time phase during each Habituation trial. A camera documented looking time and a coder blind to the trial type (i.e., helping or hindering trial) coded looking time towards the stationary shapes in real time using looking time habituation software, Habit 2 (Oakes, Sperka, & Cantrell, 2015). The total looking time toward the scene during the looking time phase was computed for each habituation trial. Each looking time phase ended when the toddler looked away continuously for 3 seconds or after 30 seconds of looking had elapsed. For children who did not look towards the scene within 10 seconds of initiation of the looking time phase, the looking time phase was ended. The curtain closed to occlude the display after each Habituation trial. Habituation trials alternated between helping and hindering (counterbalanced with constraint for first trial) until the habituation criterion was reached: sum of the looking time during three consecutive looking time phase trials is less than half of the sum of the looking time on the first three looking time phase trials (max: 12 trials). Toddlers were subsequently presented two tests to measure social attribution; the PAT Looking Time Test and the PAT Choice Test. Notably, the Looking Time Test always preceded the Choice Test.

2.3.1.2 PAT Looking Time Test Phase

During the PAT Looking Time Test, toddlers were shown test puppet scenes featuring a green plain and the climber shape situated in the middle of the scene and the helping and hindering shapes both equidistant to each side of the climber (counterbalanced with constraint). See Figure 2 for a static picture of a PAT Looking Time Test scene. The helper and hinderer shapes were of the same shape and color that the child saw during the PAT Habituation Phase. The scenes were shown on a puppet stage display made of wood (33" x 24.5") and toddlers continued to sit 3.5 feet away from the puppet stage.

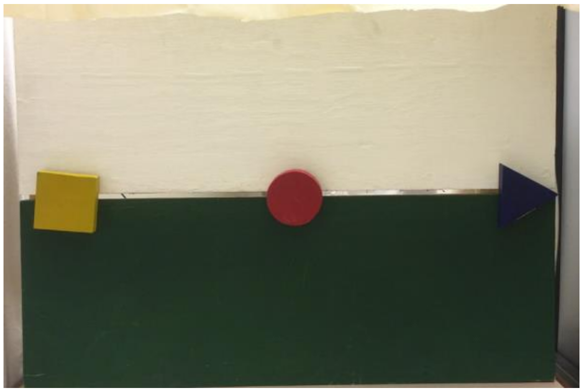

Figure 2. Static picture of a PAT Looking Time Test scene.

Toddlers were presented with six alternating helping and hindering Looking Time Test trials to measure evidence of social attribution (counterbalanced with constraint for first trial). Each Looking Time Test trial included a scene presentation phase and a looking time phase. At the start of each Looking Time Test trial, the curtain was removed to show the scene, which began the scene presentation phase. In the helping test trials, the climber wiggled then moved towards the helping shape, whereas in the hindering test trials, the climber wiggled then moved towards the hindering shape. Each scene presentation lasted approximately three seconds in duration.

A bell rang to end the scene presentation phase and begin the looking time phase. A camera documented looking time and a coder blind to test trial type (i.e., helping or hindering) coded looking time towards the stationary shapes in real time using Habit 2. The total looking time toward the scene during the looking time phase was computed for each Looking Time Test trial. Each looking time phase ended when the toddler looked away continuously for 3 seconds or when 30 seconds of looking had elapsed. For children who did not look towards the scene within 10 seconds of initiation of the looking time phase, the Looking Time Test trial was ended. Notably, these are the same criteria used in the Habituation Phase. The curtain closed to occlude the display after each Looking Time Test trial. Compliance with the Looking Time Test was defined as looking towards at least one helping and one hindering test scene during the Looking Time Test.

2.3.1.3 PAT Choice Test Phase

During the PAT Choice Test, parents were instructed to close their eyes for the task duration such as not to influence the child's behavior during this test. Toddlers were presented both the helping and hindering puppet shapes, which were placed on the table in front of them 17 inches apart (side counterbalanced with constraint). Toddlers were immediately instructed to "Touch a shape." Their choice was defined as the shape they touched first. Notably, the curtain occluded the puppet display such that the toddler could not reference the scene in making their choice. In the case that the child touched both shapes at the same time, their hands were removed from the shapes and they were instructed to "Touch *one*." This was attempted until the child clearly touched one shape before the other. Compliance with the Choice Test was defined as the child clearly touching one shape before the other. If a child refused to touch one shape after several prompts, they were considered non-compliant with the Choice Test.

2.3.2 Touchscreen Attribution Task (TAT)

The TAT is a novel, computerized adaptation of the PAT presented on a touchscreen monitor. The task also includes a Habituation Phase as well as two test phases including the Looking Time Test Phase and the Choice Test Phase. Performance on both test phases are intended to demonstrate whether children understood to social context of the scenes presented during the Habituation Phase and are considered primary measures of social attribution in the current study.

2.3.2.1 TAT Habituation Phase

Toddlers were habituated to helping and hindering social cartoons featuring a green hill and three colored shape characters (circle, square, and triangle) displayed on a touchscreen computer monitor. See Figure 3 for a static screen shot of a TAT Habituation Phase cartoon. Toddlers were seated 24 inches away from a touchscreen computer monitor (20’’x13’’). All Habituation trial stimuli (i.e., all scene proportions, colors, shape movements) and procedures were identical to the Habituation trials in the PAT with the exception of presentation modality (i.e., computer touchscreen instead of a puppet stage). Furthermore, instead of a curtain opening and closing between Habituation trials, a cross hair was displayed for two seconds along with the sound of spoons “clinking” between TAT Habituation trials. The total looking time toward the cartoon during the looking time phase was also computed for each Habituation trial during the TAT. Toddlers were then presented two tests to measure social attribution; the TAT Looking Time Test and the TAT Choice Test. Notably, the Looking Time Test always preceded the Choice Test.

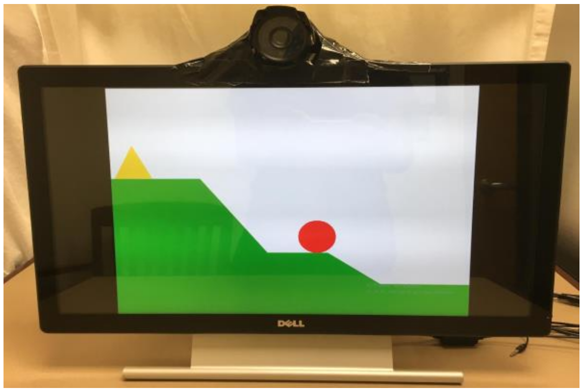

Figure 3. Static picture of a TAT Habituation Phase scene.

2.3.2.2 TAT Looking Time Test Phase

The TAT Looking Time Test is a computerized adaptation of the PAT Looking Time Test presented on a touchscreen monitor. Toddlers were shown test cartoon scenes featuring a green plain and the climber shape situated in the middle of the scene and the helping and hindering shapes both equidistant to each side of the climber (counterbalanced with constraint). See Figure 4 for a static screen shot of a TAT Looking Time Test cartoon. The helper and hinderer shapes were of the same shape and color that the child saw during the TAT Habituation Phase. The scenes were shown on the same touchscreen monitor as in the TAT Habituation Phase and toddlers continued to sit 24 inches away from the monitor. All Looking Time Test trial stimuli (i.e., all scene proportions and angles and colors) and procedures were identical to the Looking Time Test trials in the PAT with the exception of presentation modality (i.e., computer touchscreen instead of a puppet stage). The total looking time toward the cartoon during the looking time phase was also computed for each Looking Time Test trial during the TAT. Furthermore, instead of a curtain opening and closing between Looking Time Test trials, a cross hair was displayed for two seconds along with the sound of spoons “clinking” between TAT Looking Time Test trials.

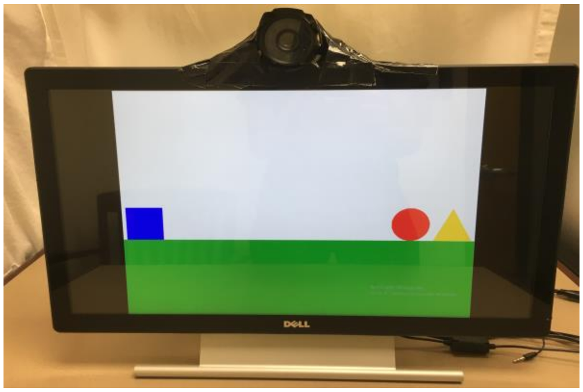

Figure 4. Static picture of a TAT Looking Time Test scene.

2.3.2.3 TAT Choice Test Phase

The TAT Choice Test was designed to be the computerized equivalent of the PAT Choice Test. During the Choice Test, the touchscreen was pulled closer to the child such that it was positioned within arm's reach for the child and centered directly in front of them. The TAT Choice Test consisted of two phases, a training phase and a choice phase. The training phase was designed to familiarize the child with the touchscreen. In the training phase, the climber shape (i.e., red circle) was displayed in the center of the screen and an audio recording was played saying, "Touch the shape." The experimenter then modeled touching the shape, which initiated a 10 second cartoon clip of the climber shape moving around the screen paired with a musical tune. The climber then appeared again in the center of the screen and an audio recording was played saying, "You touch the shape" which was meant to encourage the child to touch the shape. After the child touched the shape, the same cartoon clip of the climber shaped paired with the musical tune was initiated, which concluded the training phase. If the child did not immediately touch the climber shape, the experimenter encouraged the child to touch the shape. If the child continued to not touch the shape despite prompts, the experimenter provided hand-over-hand support for the child to touch the shape (whether the child required hand-over-hand guidance during training was recorded).

The choice phase began immediately after the training phase. The choice phase included presentation of the helping and hindering shapes presented equidistant from each other on the right and left side of the screen and the computer simultaneously played an audio recording saying, “You touch a shape.” The child’s choice was recorded as the shape they touched first, which initiated a 10 second cartoon clip of the shape they touched moving around the screen paired with a musical tune. Unlike the PAT, there were no cases where the child attempted to touch both shapes at the same time. If a child refused to touch one shape after several prompts, they were considered “non-compliant” with the task and no choice was recorded.

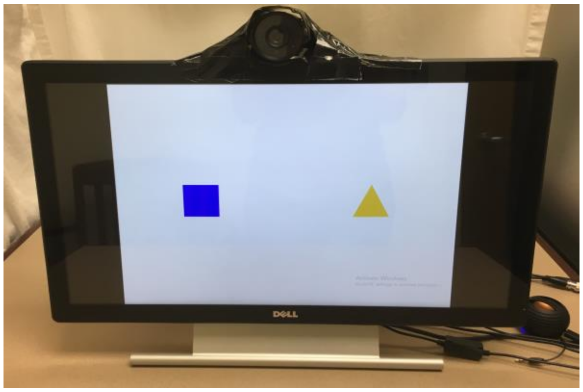

Figure 5. Static picture of the TAT Choice Test scene.

2.3.3 Reliability of looking time data

Another coder blind to trial type (i.e., helping or hindering) rescored 26% randomly-chosen trials (five toddlers’ PAT and TAT within each diagnostic group) from video using Habit 2 to establish reliability. A high degree of reliability was found across looking time coders for both experimental tasks. For the PAT, the average measure intraclass correlation (ICC) for the

Habituation Phase was .972 with a 95% confidence interval of .960 to .980 (*F*(125, 125) = 35.336, *p* < .001, and for the Looking Time Test Phase was .944 with a 95% confidence interval of .915 to .963 (*F*(89, 89) = 17.824, *p* < .001. For the TAT, the average measure ICC for the Habituation Phase was .925 with a 95% confidence interval of .895 to .946 (*F*(137, 137) = 13.283, *p* < .001 and for the Looking Time Test Phase was .856 with a confidence interval of .780-.906 (*F*(85, 85) = 6.968, *p* < .001.

2.4 Additional measures

2.4.1 The Modified Checklist for Autism in Toddlers Revised (M-CHAT-R) and Follow-Up (M-CHAT-R FU)

The M-CHAT-R/F (Robins, Fein, & Barton, 2009) is an early screening tool used to detect children who may be at risk for an ASD in low-risk samples (see Appendix 1). The M-CHAT-R is a twenty item, *yes/no* questionnaire filled out by parents, often during pediatric well-visits at 18 and 24 months. Items on the M-CHAT-R assess both typical childhood behavior (e.g., *looking toward an object a caregiver is looking at)* as well as behaviors that are commonly observed in toddlers with ASD (e.g*., unusual finger movements near the face*). Item responses are scored as either Pass or Fail and the total number of items failed is the M-CHAT Total Score. Children with a M-CHAT-R Total Score of 8 or higher are considered a screen positive (i.e., at risk for ASD) and are immediately referred for a comprehensive diagnostic evaluation. Children with a M-CHAT-R Total Score between 3 and 7 are administered the M-CHAT-R Follow-Up, to confirm appropriate item interpretation. The Follow-Up may be administered via interview or computer. Item responses are scored as either pass or fail and the total number of items failed is the M-CHAT-R

Follow-Up Total Score. Children with a M-CHAT-R Follow-Up Total Score between 2 and 7 are considered a screen positive (i.e., at risk for ASD) and are referred for a comprehensive diagnostic evaluation. Initial psychometric properties of the M-CHAT-R/F were reported in a sample of 16,115 low-risk toddlers (Robins et al., 2014). Consistency across the M-CHAT-R and Follow-up was adequate (Cronbach's alpha = 0.79). Sensitivity and specificity were estimated to be 0.854 and 0.993, respectively, and positive predictive value (PPV) was reported to be 0.475. Notably, PPV for any developmental delay or concern was 0.946.

M-CHAT-R/F Final Total Score (FTS) was used for the purposes of the current study. For children who did not screen positive on the M-CHAT-R (i.e., M-CHAT-R Total Score of 2 or less), the M-CHAT-R/F FTS was the M-CHAT-R Total Score. For children who screened positive on the M-CHAT-R, the M-CHAT-R/F FTS is the M-CHAT-R Follow-Up Total Score.

2.4.2 Autism Diagnostic Observation Schedule, Toddler Module (ADOS-2)

The ADOS-2, Toddler Module (Lord et al., 2012) is an observational measure of ASD behavior for toddlers between the ages of 12 and 30 months. The ADOS-2 is administered as a semi-structured play session structured to elicit opportunities for the individual to respond to social presses and can be administered in thirty to sixty minutes. Behaviors within two domains are assessed including Social Affect (SA) and Restricted and Repetitive Behavior (RRB). Items are scored on a 3 to 4-point scale (depending on the item), with 0 indicating *no abnormality* and 3 indicating *moderate to severe abnormality*. Domain algorithm scores are computed based on an algorithm (only items shown to best predict a diagnosis are included, and scores of 3 are converted to 2), and a total algorithm score is computed using domain algorithm scores from the SA and RRB domain domains. The Toddler Module provides two algorithms: One for children between 12 and 20 months/children aged 21 to 30 months with limited words (All Younger/Older With Few to No Words) and the other for children within the ages of 12 to 30 months and some words (Older With Some Words). The ADOS-2, Toddler Module classifies individuals into one of three categories; Little-to-No Concern, Mild-to-Moderate concern, and Moderate-to-Severe concern. Interrater reliability and test-retest reliability on the ADOS-2 Toddler Module ranged from .84 to .99 and .60 to .95, respectively, for the SA and RRB domain scores and total scores (intraclass correlations). Sensitivity ranged from 83% to 91% and specificity ranged from 86-94% for the SA and RRB domain scores and total scores. Interrater agreement in broad diagnostic classification ranged from 87% to 97%.

The ADOS-2 includes a Calibrated Severity Score (CSS) that can be used as a measure of symptom severity (Gotham, Pickles, & Lord, 2009). Scores range from 1-10, with higher scores representing greater severity. Total CSS on the ADOS-2 was used as a measure of ASD severity in the current study.

2.4.3 Mullen Scales of Early Learning (MSEL)

The MSEL (Mullen, 1995) is a direct assessment of cognitive, language and motor functioning in children from birth to 68 months. The MSEL provides T-Scores (mean of 50, standard deviation of 10) and for five scales including Visual Reception, Fine Motor, Gross Motor, Expressive Language and Receptive Language, as well as an Early Learning Composite (ELC) based on all scales except Gross Motor. Age equivalents are also provided for the five scales. Internal consistency ranging from .75 to .83 (alpha coefficients) within each scale, test-retest reliability from .75 to .96, and inter-rater reliability from .91 to .99. The ELC has shown adequate concurrent validity with the Bayley Scales of Infant Development Mental Development Index (*r* = .70; Bayley, 1969).

Standardized scores on measures such as the MSEL tend to be insensitive to variability in lower functioning toddler populations due to floor effects (Carter et al., 2007). For example, toddlers with developmental delays often bottom-out on the MSEL with a T-score of 20. In order to capture variability in item performance in developmentally delayed samples, previous studies have utilized age-equivalent (AE) scores instead of standard scores in analyses (Carter et al., 2007; Hartley & Sikora, 2009). Furthermore, given the variability in the onset of language in toddlers, a nonverbal reasoning measure, such as the MSEL Visual Reception scale score, is often used as an estimate of IQ in this age group (Carter et al., 2007; Hartley & Sikora, 2009). Therefore, a nonverbal Intelligence Quotient (NVIQ) score was computed for each subject by dividing the MSEL Visual Reception AE, by chronological age, and multiplying this number by 100. This NVIQ score was used as an estimate of IQ in this study.

2.4.4 Screen Media Questionnaire (SMQ)

The SMQ is a novel parent-report measure of screen media use in toddlers (Ludwig, Williamson, Adamson, Brosnan, Jones & Robins, 2014; see Appendix 2). The questionnaire is designed to take less than 10 minutes to complete. The questionnaire gathers information about average total screen media use on weekends and weekdays in minutes as well as more specific information about average total screen media use on weekdays and weekends in minutes, level of child engagement (5 point scale), and how the child uses the screen media (parents check off provided examples of use) for televisions, stationary computer/laptop, tablet, smart phone, and any other hand-held device with a screen. This questionnaire was developed to explore the potential impact of experience with screen media and performance on the TAT. The SMQ is a novel measure and psychometric properties are currently unavailable.

An SMQ Average Total Minutes Per Day (SMQ-MD) was computed for study analyses. The SMQ-MD was computed by multiplying the reported average number of minutes of total screen media use per weekday by five and multiplying the reported average number of minutes of total screen media use per weekend day by two and summing these totals then dividing by seven to yield the average reported minutes of screen media use per day. A similar approach was taken to compute average use of tablet and smart phone per day (SMQ Average Total Minutes Per Day for Tablet and Smart Phone; SMQ-MDTP) given that these technologies are most similar to the TAT paradigm.

2.4.5 Data analysis

2.4.5.1 Statistical analyses

Means, standard deviations and ranges or frequencies and percent of sample were computed for all variables of interest. Correlations between sample characteristics and all outcome variables of interest were examined in the whole sample. Pearson's correlation was used for continuous dependent variables and point-biserial correlation was used for dichotomous variables.

Aim 1. Aim 1 explored whether toddlers demonstrated habituation to helping and hindering scenes during the Habituation Phase on the PAT and TAT and whether habituation characteristics differed by diagnostic group. (1a) To examine whether there were differences in whether toddlers habituated to the stimuli during the PAT and the TAT, a 2x2 chi-square analysis was conducted including modality (i.e., PAT or TAT) and whether habituation criterion was met during the Habituation Phase. Two 3x2 chi-square analyses were also conducted including diagnostic group (i.e. ASD, DD and TD) and whether habituation criterion was met during the Habituation Phase. (1b) To examine whether there were differences in the number of trials required for habituation by task and diagnosis, a mixed factorial ANOVA was conducted with diagnostic group as the between subjects factor and mean number of trials presented before the habituation criterion was reached on the PAT and the TAT as the repeated factor. Notably, for the purposes of calculating means and standard deviations, toddlers who did reach habituation criterion were assigned the maximum trial count of 12. (1c) To examine whether there were differences in engagement during habituation by task and diagnosis, a mixed factorial ANOVA was conducted with diagnostic group as the between subjects factor and mean looking time per habituation trial as on the PAT and TAT as the repeated factor.

Aim 2. Aim 2 explored performance on the PAT social attribution and social evaluation measures. In order to explore whether performance on the PAT differed based on diagnosis and whether performance predicts a diagnosis of ASD and severity of ASD symptoms, analyses were conducted for both the social attribution (i.e., Looking Time Test) and social evaluation (i.e., Choice Test) measures. (2a.) To measure whether there were differences in looking time towards the helping and hindering test scenes during the Looking Time Test by diagnosis, a mixed factorial ANOVA was conducted with diagnostic group as the between subjects factor and average looking time during helping and hindering test trials as the repeated factor. (2b.) To measure whether looking time predicts a diagnosis of ASD, a logistic regression was conducted with difference in mean looking time between hindering and helping test trials (i.e., the average looking time during hindering test trials minus the average looking time during helping test trials) as the predictor and ASD status (i.e., ASD vs. non-ASD) as the outcome. (2c.) To measure whether looking time predicts ASD symptom severity, a linear regression was conducted with the difference in mean looking time between hindering and helping test trials as the predictor and ADOS-2 CSS as the outcome. (2d.) To measure whether there were group differences in shape choice on the Choice Test, a 3x2 chi-square analysis was conducted including diagnostic group by shape choice. (2e.) Additionally, to explore whether shape preference predicts an ASD diagnosis, a likelihood ratio was be computed with shape choice as the predictor and ASD status as the outcome. (2f) To examine whether ASD symptom severity differed based on shape preference, a *t*-test was conducted with shape preference as the independent variable and ADOS-2 CSS as the dependent variable.

Aim 3. Aim 3 explored performance on the TAT social attribution and social evaluation measures. In order to explore whether performance on social attribution and social evaluation measures on the TAT differed based on diagnostic group and whether performance predicts an ASD diagnosis and severity of symptoms, the same analyses from Aim 1 were conducted using dependent variables derived from the TAT (i.e., 3a. through 3f.).

Aim 4. Aim 4 served to explore potential differences in performance on measures of social attribution and social evaluation by task modality and diagnostic group. (4a.) To examine potential differences in performance on social attribution measures (i.e., Looking Time Test) by modality and diagnostic group, a mixed factorial ANOVA was conducted with diagnostic group as the between factor and difference in mean looking time between hindering and helping test trials on the PAT and the TAT as the repeated factor. (4b.) For the Choice Test, a 2x2x3 loglinear analysis was conducted including shape preference on the PAT and TAT by diagnostic group.

2.4.5.2 Power analysis

A study utilizing a similar social attribution paradigm in infants demonstrated a medium size effect (*f* = .25; Kuhlmeier et al., 2003). A power analysis (G-Power V3.1) indicated that 42 toddlers would provide sufficient power (1 – β = .95) at α = .05 to detect a medium sized effect in a 3x2 mixed factorial ANOVA, and that 55 toddlers would provide sufficient power (1 – β = .80) at α = .05 to detect a medium sized effect in a linear regression analysis.

2.4.6 A note about whether to include NVIQ as a covariate in analyses

Volkmar and colleagues (1993) proposed that deciding whether it is appropriate to control for cognitive ability/intellectual functioning requires theoretical justification. The authors suggest that control for IQ is appropriate if IQ is conceptualized as the cause of behavioral differences, but would be inappropriate if it is an associated feature or result of behavioral differences. Controlling for IQ in the latter scenario might lead to, "the control of factors that are not confounding variables," (Volkmar et al., 1993, p. 581). Dennis and colleagues (2009) put forth strong arguments against the conceptualization of IQ as causal and use of IQ as a covariate in the study of any neurodevelopmental condition. The authors have suggested that IQ does not predate, but "postdates the [neurodevelopmental] condition, charts the history of the condition, is always confounded with and/or by the condition, and can never be separated from the effects of the condition" (p. 2). The authors provide theoretical evidence against the idea that intelligence represents an innate construct that predicts aptitude and potential, but rather represents a history of achievement and performance shaped by the condition. According to Dennis and colleagues, this common misconception has led to the idea that IQ is a contributing factor to disorder and should be controlled when other causal variables are of interest. Given these compelling arguments, NVIQ will not be included as a covariate for subsequent analyses.

3 RESULTS

First, similarities and differences in the sample characteristics in the three groups of children are presented in Section 3.1. This is important for determining whether there are any differences in the samples that may explain habituation characteristics and performance on measures of social attribution by diagnostic group. Second, information about rates of task compliance and participant-specific task variables including seating arrangement and rates of hand-over-hand assistance during touchscreen training, which are also important variables to consider in interpreting task performance, are presented in Section 3.2. Finally, results of the primary research questions are presented in Section 3.3.

3.1 Sample characteristics

Means, standard deviations, ranges, and frequencies for all sample characteristics can be found in Table 2. All significant correlations are reported.

Table 2. *Characteristics of Toddlers Included in the Final Sample*

Variable	Total (*n* = 57)	ASD (*n* = 15)	DD (*n* = 20)	TD (*n* = 22)
Age at Session 1 (months)				
Mean	21.61	22.31	21.15	21.54
Range	15-29	19-29	15-27	18-27
Standard Deviation	2.98	2.90	3.04	3.01
Age at Session 2 (months)				
Mean	22.07	22.69	21.51	22.15
Range	16-29	19-29	16-28	18-27
Standard Deviation	2.97	2.76	3.15	2.98
Days Between Sessions				
Mean	14.04	11.60	10.90	18.55
Median	7.00	5.00	7.00	15.50
Range	1-84	1-84	1-28	1-51

Standard Deviation	14.60	20.68	8.63	13.44
Sex				
Male (%)	34 (59.6)	12 (80.0)	12 (60.0)	10 (45.5)
Female (%)	23 (40.4)	3 (20.0)	8 (40.0)	12 (54.5)
MSEL VR AE (months)				
Mean	17.49	15.60	14.55	21.45
Range	5-33	6-25	5-20	15-33
Standard Deviation	5.79	5.21	4.24	5.28
NVIQ				
Mean	82.02	70.91	70.22	100.32
Range	21.48-151.10	22.16-119.67	21.48-116.81	67.03-151.10
Standard Deviation	27.33	24.48	22.87	23.39
M-CHAT-R/F FTS	(*n* = 56)	(*n* = 15)	(*n* = 20)	(*n* = 20)
Mean	4.33	6.13	5.00	.30
Range	0-19	2-19	1-15	0-2
Standard Deviation	4.37	4.75	3.99	0.67
ADOS-2 CSS	(*n* = 34)	(*n* = 15)	(*n* = 19)	-
Mean	5.21	7.93	3.05	-
Range	1-10	2-10	1-8	-
Standard Deviation	3.07	1.98	1.78	-
SMQ-MD	(*n* = 51)	(*n* = 13)	(*n* = 16)	(*n* = 22)
Mean	73.86	131.81	60.89	49.05
Range	0.00-300.00	10.00-300.00	0.00-240.00	0.00-137.14
Standard Deviation	9.91	91.47	66.40	33.65
SMQ-MDTP	(*n* = 52)	(*n* = 13)	(*n* = 17)	(*n* = 22)
Mean	21.37	28.87	14.29	22.41
Range	0.00-91.43	0.00-68.57	0.00-60.00	0.00-91.43
Standard Deviation	21.41	21.72	17.102	23.34

Note. ADOS-2 CSS = Autism Diagnostic Observation Schedule-2 Calibrated Severity Score (Range = 1-10); M-CHAT-R/FU FTS = M-CHAT-R/FU Final Total Score (Range = 0-20); MSEL VR AE = Mullen Scales of Early Learning Visual Reception Age Equivalent; NVIQ = Nonverbal Intelligence Quotient (Standard Score); SMQ-MD = Screen Media Questionnaire Average Total Minutes Per Day Total; SMQ-MDTP = Screen Media Questionnaire Average Total Minutes Per Day for Tablet and Smart Phone.

3.1.1 Age

A one-way between subjects ANOVA revealed no significant difference in age at Session 1 ($F(2, 56) = .648$, $p = .527$, $\eta^2 = .023$) or Session 2 ($F(2, 56) = .683$, $p = .509$, $\eta^2 = .024$) or in the number of days between sessions ($F(2, 56) = 1.768$, $p = .180$, $\eta^2 = .061$) across diagnostic groups. Age was correlated with the difference in looking time toward hindering and helping trials during the Looking Time Test for the PAT only ($r = .278$, $p = .036$) and was therefore included as a covariate in analyses utilizing this variable. Notably, this correlation was no longer significant when toddlers with MSEL AE less than 12 months were removed from analyses ($r = 185$, $p = 193$). Age was not significantly correlated with any other outcome variables of interest.

3.1.2 NVIQ

A one-way between subjects ANOVA demonstrated a significant difference in NVIQ across groups ($F(2, 56) = 10.871$, $p < .001$, $\eta^2 = .29$). As expected, a Gabriel's procedure post hoc test (used due to differences in sample size) revealed that NVIQ was significantly lower in the ASD ($p = .001$) and DD ($p < .001$) groups compared to the TD group. There was no statistically significant difference in NVIQ between the ASD and DD groups ($p > .999$). Notably, there were four toddlers in the TD group with NVIQ scores below a standard score of 80. Upon examination, three of these children had overall MSEL Early Learning Composite (ELC) scores in the average range. One toddler had an ELC in the borderline range of functioning; however clinical observation indicated significant compliance issues during administration of the MSEL and therefore this child was also included in analyses. Overall, it was decided that all four children would be included in the TD group given no indication of developmental delay and to preserve natural variance.

Notably, research has shown that children less than 12 months require increased scaffolding (i.e., eyes on the shapes) during social attribution tasks (Hamlin et al., 2007); however, six children included in the final sample had an AE lower than 12 months ($n_{\text{ASD}} = 2$, $n_{\text{GDD}} = 4$). In order to preserve power, all reported analyses were conducted including these children given that removing them from analyses resulted in no changes to the significance of any of the reported analyses nor meaningful changes to effect sizes, unless otherwise stated in the sections below.

NVIQ was correlated with mean looking time per trial during habituation (*r* = .325, *p* = .014) and choice during the Choice Test (choice of helper coded as 1; *r* = .270, *p* = .046) on the PAT in the whole sample. Nonetheless, NVIQ was not included as a covariate in analyses given theoretical arguments against this practice (Dennis et al., 2009; Volkmar et al., 1993) and because it did not meet statistical assumptions for inclusion as a covariate (i.e., differences in NVIQ across diagnostic groups). When children with MSEL AE less than 12 months were removed from analyses, NVIQ continued to correlate with mean looking time per trial during habituation (*r* = .334, *p* = 193), but the relationship between NVIQ and choice during the Choice Test was no longer significant (*r* = .181, *p* = .213). NVIQ was not correlated with any other outcome variables of interest.

3.1.3 M-CHAT-R/F and ADOS-2

As expected, variance in final scores on the M-CHAT-R/F differed between diagnostic groups since scores in the TD are restricted to 0-2 (Levine's Test: *F*(2, 52) = 11.202, *p* < .001). As such, Welch's *F* is reported to account for violation of homogeneity of variance. A one-way between subjects ANOVA demonstrated a significant difference in M-CHAT-R/F Final Total Score across groups (*F*(2, 22.123) = 23.74, *p* < .001, η^2 = .362). As expected, a Games-Howell post hoc test (used due to unequal variances) revealed that the M-CHAT-R/F Final Total Score was significantly higher in the in the ASD (*p* = .001) and DD (*p* <.001) groups compared to the TD group. There was no statistically significant difference in M-CHAT-R/F Final Total Score between the ASD and DD groups (*p* = .738). Notably, there were two children in the TD group with missing M-CHAT-R/F scores.

Also as expected, an independent samples *t*-test revealed a significant difference in ADOS-2 CSS between the ASD and DD groups (*t* = -7.557, *p*<.001, Cohen's *d* = 2.59). Notably, the ADOS-2 score was unavailable for one child within the DD group as the score was considered invalid due to severely impaired cognitive abilities (i.e., MSEL Visual Reception AE = 5 months).

M-CHAT-R/F Final Total Scores and ADOS-2 CSS were not correlated with any outcome variables of interest.

3.1.4 SMQ-MD & SMQ-MDTP

One participant's (DD) SMQ-MD and SMQ-MDTP scores were considered extreme outliers (i.e., three times the interquartile range; SMQ-MD = 411.43, SMQ-MDTP = 385.71) and these scores were removed from analyses. Also, one parent did not provide data for the overall screen media use (i.e., SMQ-MD), but did provide data for tablet and phone use such that the SMQ-MDTP could be computed. Parents of four participants did not return the SMQ; these data are missing.

Levine's Test indicated violation of homogeneity of variance for SMQ-MD scores across diagnostic groups (*F*(2, 48) = 5.624, *p* = .006). As such, Welch's *F* is reported. A one-way between subjects ANOVA demonstrated a significant difference in SMQ-MD across groups *(F*(2, 21.56) = 4.829, *p* = .018, η^2 = .239). A Games-Howell post hoc test (used due to unequal variances) revealed that SMQ-MD was significantly higher in the ASD group compared to the TD group (*p* = .019) and this approached significance in the DD group (*p* = .072; when run with a Bonferroni post hoc test ASD vs. DD was significant at *p* = .012). Considering this, it is not surprising that SMQ-MD was also correlated with ADOS-2 CSS (*r* = .464, *p* = .013). SMQ-MD was also correlated with NVIQ (*r* = -.280, *p* = .047) in the whole sample. Notably, parents reported no use of screen time of any kind for three toddlers ($n_{\mathrm{DD}} = 2$, $n_{\mathrm{TD}} = 1$).

A one-way between subjects ANOVA demonstrated no difference in SMQ-MDTP across groups *F*(2, 51) = 1.808, *p* = .175, $\eta^2 = .069$). Notably, parents reported no use of tablet or phone for seven toddlers in the sample ($n_{\mathrm{ASD}} = 1$, $n_{\mathrm{DD}} = 4$, $n_{\mathrm{TD}} = 2$). SMQ-MD and SMQ MDTP were not correlated with any outcome variables of interest.

3.2 Compliance and participant-specific task variables

As previously mentioned, three children did not tolerate the task during the Habituation Phase and were considered "fuss outs" and these children were not included in the final sample. Because of the small number of children who fussed out, a chi-square test could not be computed given the small number of cases within the "fuss out" category for each diagnosis.

Although habituation data was collected for the 57 toddlers who were included in the final sample, some of these toddlers did not meet compliance criteria for the PAT or TAT Looking Time Test or Choice Tests. Toddlers who did not meet criteria for compliance for a test were not included in analyses examining that specific test. Table 3 provides information about the number of toddlers who did not meet compliance criteria on each test by diagnostic group. Notably, point-biserial correlations indicated no relationship between SMQ-MD or SMQ-MDTP and compliance on the PAT or TAT Looking Time Test or Choice Test (*p*s > .05); however these findings should be interpreted with caution given the very small number of children who met criteria for non-compliance on each test.

Table 3. *Non-Compliance Rates on PAT and TAT Looking Time Test and Choice Test in the Final Sample*

Test	Total (*n* = 57)	ASD (*n* = 15)	DD (*n* = 20)	TD (*n* = 22)
Choice Test				
PAT Choice Test *n* (%)	2 (3.5)	1 (6.7)	0 (0.0)	1 (4.5)
TAT Choice Test *n* (%)	6 (10.5)	2 (13.3)	3 (15)	1 (4.5)
Looking Time Test				
PAT Looking Time Test *n* (%)	2 (3.5)	1 (6.7)	1 (5.0)	0 (0.0)
TAT Looking Time Test *n* (%)	3 (5.3)	0 (0.0)	1 (5.0)	2 (9.1)

To measure whether there were group differences in seat context (i.e., parent lap or highchair) during the PAT or TAT, two 3x2 chi-square analyses were conducted including diagnostic group and seat context. Results indicated no significant difference in seat context by diagnosis for the PAT ($X^2(2, N = 57) = .029, p = .986$) or the TAT ($X^2(2, N = 57) = 1.016, p = .602$). Table 4 provides characteristics about parent seat choice.

Table 4. *Seat and Training Characteristics in the Final Sample*

Variable	Total (*n* = 57)	ASD (*n* = 15)	DD (*n* = 20)	TD (*n* = 22)
Parent Seat Choice PAT				
Parent Lap *n* (%)	42 (73.7)	11 (73.3)	15 (75.0)	16 (72.7)
Highchair *n* (%)	15 (26.3)	4 (26.7)	5 (25.0)	6 (27.3)
Parent Seat Choice TAT				
Parent Lap *n* (%)	38 (66.7)	9 (60.0)	15 (75.0)	14 (63.6)
Highchair *n* (%)	19 (33.3)	6 (40.0)	5 (25.0)	8 (36.4)
TAT Choice Test Training	(*n* = 56)	(*n* = 15)	(*n* = 20)	(*n* = 21)
Touched shape on own *n* (%)	42 (73.7)	12 (80.0)	13 (65.0)	17 (81.0)
Hand-over-hand *n* (%)	14 (24.6)	3 (20.0)	7 (35.0)	4 (19.0)

To measure whether there were group differences in whether hand-over-hand assistance was needed during TAT Choice Test training, a 3x2 chi-square analysis was conducted including diagnostic group and training type (i.e., on own or hand-over-hand). Results indicated no significant difference in training type by diagnosis ($X^2(2, N = 56) = 1.663, p=.435$). Choice training data were missing for one child in the TD group. Table 4 provides information about rates of toddlers requiring hand-over-hand training. Notably, point-biserial correlations indicated no relationship between SMQ-MD or SMQ-MDTP and whether hand-over-hand training was required (*p*s > .05) in the whole sample and by diagnosis; however this relationship approached significance in the ASD group (*r* = -.489, *p* = .090).

3.3 Results of primary aims

3.3.1 Aim 1 results: Habituation

Data from the Habituation Phase for the PAT and TAT are represented in Table 5.

Table 5. *Habituation Characteristics for the PAT and TAT in the Final Sample*

Variable	Total (*n* = 57)	ASD (*n* = 15)	DD (*n* = 20)	TD (*n* = 22)
PAT Habituation Phase				
Number of Trials				
Mean	9.26	9.20	10.05	8.59
Standard Deviation	2.62	2.89	2.44	2.51
n Habituated in 6 (%)	19 (33.3)	6 (40.0)	4 (20.0)	9 (40.9)
n Habituated in 9 (%)	14 (24.6)	2 (13.3)	5 (25.0)	7 (31.8)
n Habituated in 12 (%)	11 (19.3)	3 (5.0)	6 (30.0)	2 (9.1)
n Not Habituated (%)	13 (22.8)	4 (26.7)	5 (25.0)	4 (18.2)
Mean Looking Time (s)				
Mean	8.94	8.24	7.89	10.39
Median	7.75	7.53	8.00	7.59
Range	1.08-21.41	3.32-14.72	1.78-14.61	1.08-21.41
Standard Deviation	4.56	3.39	3.30	5.85
TAT Habituation Phase				
Number of Trials				
Mean	9.53	9.60	9.45	9.55
Standard Deviation	2.28	2.32	2.24	2.39
n Habituated in 6 (%)	12 (21.1)	3 (20.0)	4 (20.0)	5 (22.7)
n Habituated in 9 (%)	23 (40.4)	6 (40.0)	9 (45.0)	8 (36.4)
n Habituated in 12 (%)	4 (7.0)	0 (0.0)	3 (15.0)	1 (4.5)
n Not Habituated (%)	18 (31.6)	6 (40.0)	4 (20.0)	8 (36.4)
Mean Looking Time (s)				
Mean	8.43	8.88	6.86	9.56
Median	8.00	7.98	6.71	8.51
Range	2.46-22.17	3.89-16.99	2.46-10.88	3.63-22.17
Standard Deviation	3.84	3.37	2.80	4.57

Note. Number of Trials = Number of trials presented before habituation criterion was reached; s = seconds; Mean Looking Time = Mean looking time per habituation trial.

3.3.1.1 Aim 1a. & 1b.

No difference emerged in whether toddlers reached habituation criterion within 12 exposures or not on the PAT vs. the TAT *(*X^2(2, *N* = 57) = .563, *p* = .520) or whether habituation criterion was reached within 12 exposures or not across diagnostic groups for both the PAT (X^2(2, *N* = 57) = 1.966, *p* = .374) and the TAT (X^2(2, *N* = 57) = .449, *p* = .799). A mixed factorial ANOVA revealed no significant interaction (*F*(2, 54) = 1.087, *p* = .345, partial η^2 = .039) or main effect of diagnosis (*F*(2, 54) = 1.794, *p* = .457, partial η^2 = .029) or task modality (*F*(1, 54) = .0368, *p* = .547, partial η^2 = .007) on the number of habituation presentations required before the habituation criterion was met. Taken together, results suggests that toddlers, across diagnostic groups, are habituating at similar rates and require a similar number of trials to reach habituation criterion on the live and touchscreen versions of the task. Thus, we can be more confident that findings on experimental tests of social attribution and social evaluation (i.e., Looking Time Test and Choice Test) are not due to differences in habituation rates across diagnostic groups.

3.3.1.2 Aim 1c.

A mixed factorial ANOVA revealed no significant interaction (*F*(2, 54) = 6.832, *p* = .562, partial η^2 = .021) or main effect of task modality (*F*(1, 54) = .383, *p* = .539, partial η^2 = .007) on average looking time during each habituation trial (i.e., engagement); however a significant main effect of diagnostic group emerged (*F*(2, 54) = .0368, *p* = .047, partial η^2 = .107). Gabriel's procedure post hoc test (used due to differences in sample size) indicated that toddlers within the TD group on average tended to look significantly longer than toddlers in the DD (p = .043) group whereas toddlers in the ASD group looked an intermediate amount (ps > .05). These results are presented in Figure 6. Notably, when this analyses was conducted excluding the six children with MSEL VR <12 months, this main effect was trending towards significance no longer significant (F(2, 48) = 2.403, p = .101, partial η^2 = .091). Notably, there was no significant difference in mean looking time per habituation trial between toddlers in the DD group with MSEL VR < 12 (m = 7.04) compared to those with MSEL VR >12 (m = 8.09; t(18) = -.555, p = .586, Cohen's d = .286). Therefore, the loss of significance when these toddlers are removed is likely due to a loss of power.

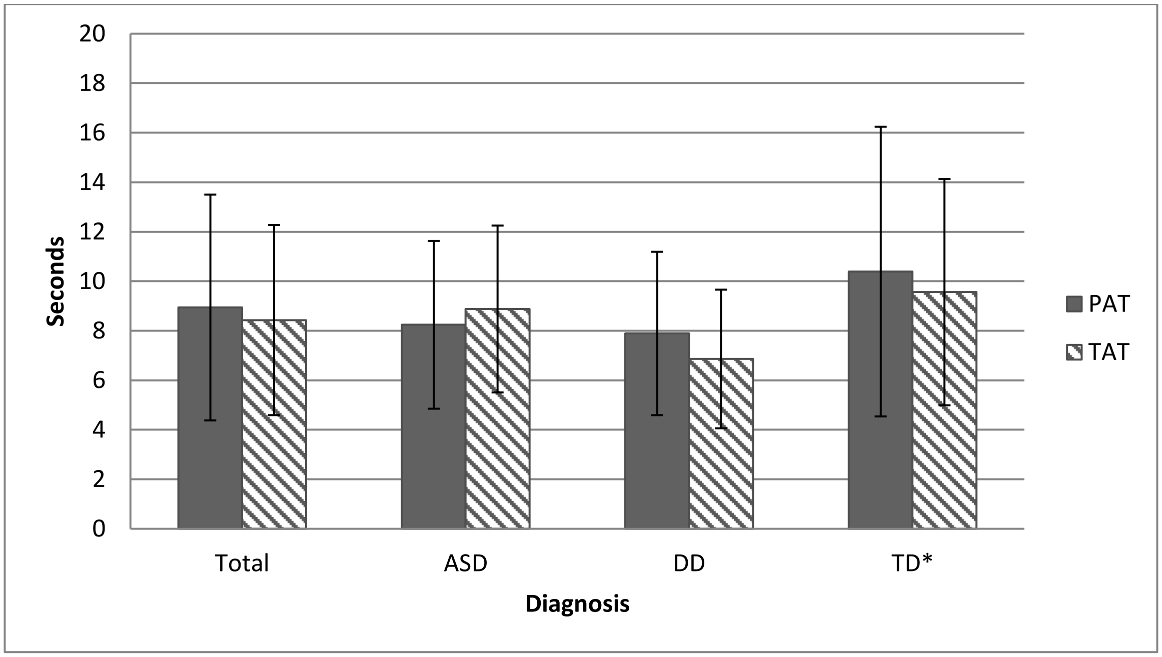

Figure 6. Mean looking time per habituation trial during the PAT and TAT.
**p* < .05 compared to DD for PAT and TAT.

3.3.2 Aim 2 results: PAT

Performance variables for the PAT are represented in Table 6.

Table 6. *Performance on the Looking Time Test and Choice Tests for PAT and TAT in the Final Sample*

Variable	Total	ASD	DD	TD
PAT Looking Time Test	(*n* = 55)	(*n* = 14)	(*n* = 19)	(*n* = 22)
Mean LT Toward Hinderer				
Mean	6.69	5.88	8.29	5.82
Range	0.70	2.70-12.23	1.27-25.20	.70-23.23
Standard Deviation	4.93	3.02	5.70	5.05
Mean LT Toward Helper				
Mean	6.12	5.36	7.00	5.84
Range	0.53-25.20	1.07-15.33	0.53-23.23	0.87-25.20
SD Standard Deviation	5.34	3.94	5.99	5.65
LT Difference				
Mean	0.57	0.52	1.29	-0.02
Range	-10.37-10.23	-8.30-4.47	-10.37-8.37	-7.66-10.23
Standard Deviation	3.90	3.06	4.84	3.52
TAT Looking Time Test	(*n* = 54)	(*n* = 15)	(*n* = 19)	(*n* = 20)
Mean LT Toward Hinderer				
Mean	5.17	6.01	4.42	5.25
Range	1.07-12.03	2.40-11.13	1.13-8.57	1.07-12.03
Standard Deviation	2.83	2.87	2.42	3.10
Mean LT Toward Helper				
Mean	5.29	5.07	5.17	5.57
Range	0.73-13.47	1.30-11.90	1.53-13.47	0.73-13.17
Standard Deviation	3.25	2.94	3.20	3.64
LT Difference				
Mean	-0.12	0.94	-0.75	-0.32
Range	-9.13-7.96	-7.75-7.77	-7.27-3.90	-9.13-7.96
Standard Deviation	3.82	3.98	3.45	4.06
PAT Choice Test	(*n* = 55)	(*n* = 14)	(*n* = 20)	(*n* = 21)
n Choosing Hinderer (%)	27 (50.9)	10 (71.4)	12 (60.0)	6 (28.6)
n Choosing Helper (%)	28 (49.1)	4 (28.6)	8 (40.0)	15 (71.4)
TAT Choice Test	(*n* = 51)	(*n* = 13)	(*n* = 17)	(*n* = 21)
n Choosing Hinderer (%)	36 (70.6)	8 (61.5)	11 (64.7)	17 (81.0)
n Choosing Helper (%)	15 (29.4)	5 (38.5)	6 (35.3)	4 (19.0)

Note. LT = Looking Time; LT Difference = Mean Looking time Toward Hinderer-Mean Looking Time Toward Helper (i.e., positive values indicate increased mean looking time toward Hinderer scenes and negative values indicate increased mean looking time towards Helper scenes); Mean looking time across the three Hinderer test trials; Mean Looking Time Toward Helper = Mean looking time across the three Hinderer test trials.

3.3.3 Aim 2a.

To measure whether there are group differences in looking time towards the helping vs. hindering shapes during the PAT Looking Time Test, a 3x2 mixed factorial ANOVA was conducted with diagnostic group (i.e., ASD, DD, or TD) as the between subjects factor, and Looking Time Test trial type (i.e., mean looking time toward helping test trials and mean looking time toward hindering test trials) as the repeated factor. The interaction *(F*(2, 52) = 0.564, *p* = .572, partial η^2 = .021), and main effects of diagnosis (*F*(2, 52) = 1.0, *p* = .375, partial η^2 = .037) and trial *(F*(1, 52) = 2.217, *p* = .275, partial η^2 = .021) were not significant.

Exploratory Follow-Up Analyses: Given the unexpected non-significant difference in mean looking time between the helping and hindering test trials for the typically developing group (Hamlin et al., 2007, 2010; Kuhlmeier et al., 2003), which has been demonstrated in a number of studies in infants and pilot data in toddlers (Gonsiorowski et al., 2015) follow-up analyses were conducted in an effort to better understand the current findings (Table 7).

First, characteristics of looking time during the first two Looking Time Test trials were examined. Kuhlmeier and colleagues (2003) reported that some children did not tolerate all six test trials given limited attention spans in infants and thus, analyses were conducted with the first two trials only, per procedures of Gergely and colleagues (1995). We replicated these procedures by conducting a 3x2 mixed factorial ANOVA with diagnostic group as the between subjects factor, and mean looking time towards the first Helping or Hindering test trial as the repeated factor. The interaction *(F*(2, 52) = 0.01, *p*=.987, partial η^2 = .001), and main effects of diagnosis (*F*(2, 52) = 0.27, *p* = .762, partial η^2 = .010) and trial *(F*(1, 52) = 0.02, *p* = .900, partial η^2 < .001) were not significant.

Second, to examine whether there were any differences in mean looking time across all six Looking Time Test Trials a mixed factorial ANOVA was conducted. Because the assumption of sphericity was rejected ($X^2(14) = 36.38$, $p = .001$) the Greenhouse-Geisser correction was used. This analysis indicated a non-significant interaction ($F(10, 421.16) = 1.39$, $p = .201$, partial $\eta^2 = .051$) and main effect of diagnosis($F(1, 52) = 1.00$, $p = .375$, partial $\eta^2 = .037$), however significant effects of trial emerged. Specifically, there was a significant linear trend ($F(1, 52) = 16.94$, $p < .001$, partial $\eta^2 = .246$) indicating that as trials progressed, looking time decreased proportionately. Furthermore, a repeated planned contrast indicated no significant difference in looking time from each successive trial, with the exception of the difference in looking time between the first and second trial ($F(1, 52) = 5.318$, $p = .025$, $\eta^2 = .093$). Table 7 presents looking time characteristics across PAT Looking Time Test trials.

Third, mean looking time towards the initial PAT Habituation trial and mean looking time towards the initial Looking Time Test trial were compared to examine whether there were significant differences in looking time towards both novel scenes to better understand whether toddlers became bored with the task. A mixed factorial ANOVA was conducted with diagnostic status as the between subjects factor and mean looking time toward the first Habituation trial and mean looking time towards the first Looking Time Test trial as the repeated factor. Results indicated a non-significant interaction ($F(2, 52) = 1.59$, $p = .213$, partial $\eta^2 = .058$) and main effect of diagnosis ($F(1, 52) = 0.98$, $p = .383$, partial $\eta^2 = .036$), however a significant effect of trial emerged ($F(1, 52) = 10.46$, $p = .002$, partial $\eta^2 = .167$) indicating that overall, toddlers looked longer towards the first Habituation trial compared to the first Looking Time Test trial.

Table 7. *Looking Time Characteristics for PAT Looking Time Test Trials by Trial Number and First Hinderer and Helper Trial for the Final Sample*

PAT Looking Time Test Trial	Total (*n* = 55)	ASD (*n* = 14)	DD (*n* = 19)	TD (*n* = 22)
Trial 1				
Mean	9.78	9.56	9.93	9.79
Range	0.20-30.00	1.60-30.00	0.20-30.00	0.30-30.00
Standard Deviation	8.30	9.19	7.93	8.42
Trial 2				
Mean	7.30	8.90	8.04	5.64
Range	0.10-30.00	1.30-30.00	0.10-24.80	0.10-15.90
Standard Deviation	6.84	7.68	7.91	5.00
Trial 3				
Mean	6.21	5.84	6.97	5.79
Range	0.10-30.00	0.10-16.70	0.10-30.00	0.10-30.00
Standard Deviation	7.62	5.67	9.36	7.31
Trial 4				
Mean	5.07	5.20	5.92	4.26
Range	0.10-30.00	0.10-12.60	0.10-30.00	0.10-30.00
Standard Deviation	6.01	4.41	6.97	6.16
Trial 5				
Mean	4.20	1.68	5.66	4.55
Range	0.01-24.80	0.10-4.90	0.01-19.70	0.10-24.80
Standard Deviation	4.88	1.49	5.35	5.40
Trial 6				
Mean	5.86	2.54	9.37	4.94
Range	0.10-30.00	0.30-13.60	0.70-30.00	0.10-30.00
Standard Deviation	7.48	3.69	9.26	6.53
First Hinder Trial				
Mean	8.47	9.12	9.03	7.58
Range	0.10-30.00	2.60-30.00	.20-24.80	0.10-30.00
Standard Deviation	7.57	7.66	7.71	7.68
First Helper Trial				
Mean	8.61	9.34	8.94	7.85
Range	0.10-30.00	1.30-30.00	0.10-30.00	0.10-30.00
Standard Deviation	7.83	9.22	8.24	6.78
First Habituation Trial				
Mean	14.81	11.52	13.84	17.73
Range	0.80-30.00	3.60-29.90	1.40-29.70	0.80-30.00
Standard Deviation	9.05	7.33	7.88	10.35

3.3.4 Aim 2b.

To measure whether looking time predicts a diagnosis of ASD, a logistic regression was conducted with age (first step) difference in looking time towards the hinderer and helper during Looking Time Test Trials as the predictor (i.e., mean looking time towards Hinderer – mean looking time towards Helper), and ASD status (i.e., ASD vs. non-ASD) as the outcome. Results indicated that age did not contribute significantly to the model (Cox and Snell $R^2 = .020$, $p = .288$.), and introducing difference in looking time was not significant (Cox and Snell $R^2 = .020$, $p = .946$).

3.3.5 Aim 2c.

To measure whether looking time predicts ASD symptom severity, a hierarchical linear regression was conducted with age (first step) and the difference in looking time towards the hinderer and helper during Looking Time Test trials as the predictors, and ADOS-2 CSS as the outcome. Results indicated that age did not contribute significantly to the model ($F(1, 32) = 0.476$, $p = .495$, $R^2 = .015$), and introducing the difference in looking time was also not significant ($F(1, 31) = .231$, $p = .795$, $R^2 = .015$); thus neither age nor difference in looking time were significant predictors of ASD symptom severity.

3.3.6 Aim 2d.

To measure whether there are group differences in shape choice on the Choice Test, a 3x2 chi-square analysis was conducted including diagnostic group and shape choice. The analysis revealed a significant difference in choice by diagnostic group ($X^2(2, N = 55) = 7.213$, $p = .027$). Specifically, children in the TD group chose the Helper significantly more often than the Hinderer; however, this was not observed in the ASD and DD groups. Results are depicted in

Figure 7. When this analyses was conducted excluding the six children with MSEL VR <12 months, this effect was trending significance (X^2(2, *N* = 49) = 7.213, *p* = .071).

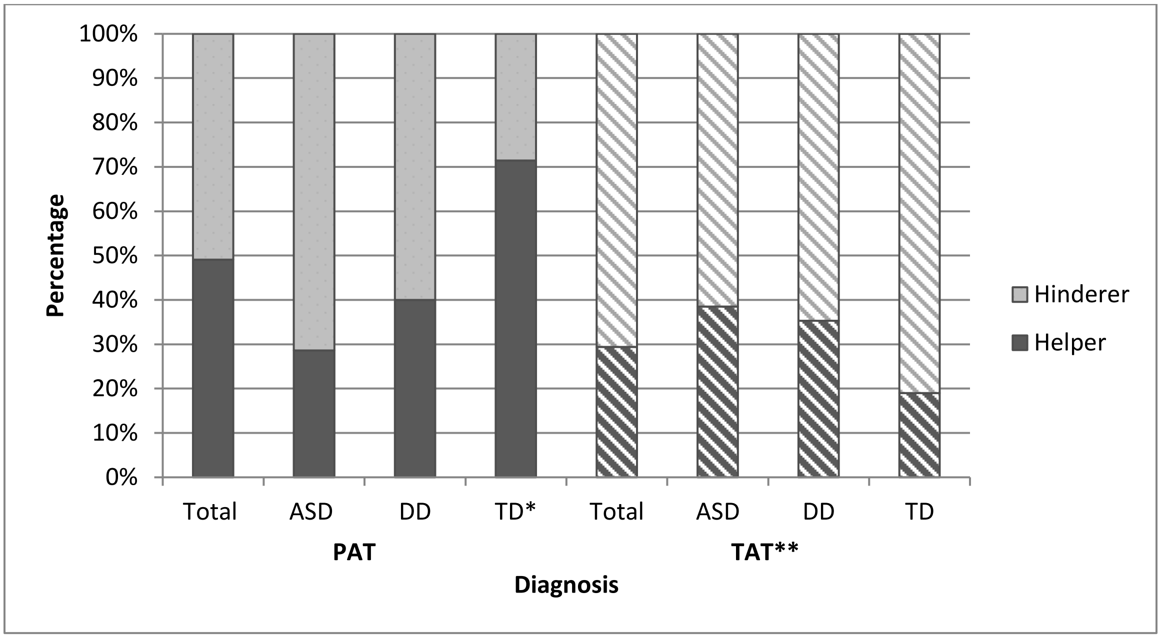

Figure 7. Percentage of toddlers choosing the Hinderer and Helper during the PAT and TAT Choice Test.
**p* = .027; ***p* = .003.

3.3.7 Aim 2e.

Additionally, to explore whether shape choice predicts an ASD diagnosis, a likelihood ratio was computed with shape choice as the predictor and ASD status as the outcome; however, this was not significant, LR(1, *N* = 55) = 3.25, *p* = .071.

3.3.8 Aim 2f.

To examine whether ASD symptom severity differs based on shape choice a *t*-test was conducted with shape choice as the independent variable and ADOS-2 CSS as the dependent variable; however, this was not significant, *t*(31) = .517, *p* = .609, Cohen's *d* = .184

3.3.9 Aim 3 results: TAT

Performance variables for the TAT are represented in Table 6.

3.3.10 Aim 3a.

To measure whether there are group differences in looking time towards the helping vs. hindering shapes during the Looking Time Test, a 3x2 mixed factorial ANOVA was conducted with diagnostic group as the between subjects factor, and Looking Time Test trial as the repeated factor. The interaction *(F*(2, 52) = 0.564, *p* = .572, partial η^2 = .021), and main effects of diagnosis (*F*(2, 52) = 1.0, *p* = .375, partial η^2 = .037) and trial *(F*(1, 52) = 2.217, *p* = .275, partial η^2 = .021) were not significant. Notably, all effects were small.

Exploratory Follow-Up Analyses: As discussed in the results section of Aim 3a, follow-up analyses were conducted in an effort to better understand these null findings for the TAT (Table 8).

First, characteristics of looking time during the first two Looking Time Test trials were examined. We replicated Kuhlmeier and colleagues (2003) procedures by conducting a 3x2 mixed factorial ANOVA with diagnostic group as the between subjects factor, and mean looking time towards the first Helping or Hindering test trial as the repeated factor. The interaction *(F*(2, 51) = 1.68, *p* = .196, partial η^2 = .062), and main effects of diagnosis (*F*(2, 51) = 0.982, *p* = .382, partial η^2 = .037) and trial *(F*(1, 51) = 0.035, *p* = .852, partial η^2 < .001) were not significant.

Second, to examine whether there were any differences in mean looking time across all six Looking Time Test Trials a mixed factorial ANOVA was conducted. Because the assumption of sphericity was rejected (X^2(14) = 41.46, *p*<.001) the Greenhouse-Geisser correction was used. This analysis indicated a non-significant interaction *(F*(7.51, 191.52) = .790, *p* = .605, partial η^2 = .030), main effect on of diagnosis *(F*(1, 51) = .493, *p* = .614, partial η^2 = .019), however a significant effect of trial emerged (*F*(3.76, 191.59) = 9.249, *p* <.001, partial η^2 = .154. Specifically, there was a significant linear trend, ($F(1, 51) = 17.140$, $p<.001$, partial $\eta^2 = .252$) indicating that as trials increased, looking time decreased proportionately. Furthermore, a repeated planned contrast indicated no significant difference in looking time from each successive trial, with the exception of the difference in looking time between the first and second trial ($F(1, 51) = 9.93$, $p = .003$, partial $\eta^2 = .163$), and the second and third trial ($F(1, 51) = 8.51$, $p = .005$, partial $\eta^2 = .143$).

Third, mean looking time towards the initial TAT Habituation trial and mean looking time towards the initial Looking Time Test trial were compared following procedures conducted by Webb and colleagues (2010) in their follow-up examination of unexpected null findings. A mixed factorial ANOVA was conducted with diagnostic status as the between subjects factor and mean looking time toward the first Habituation trial and mean looking time towards the first Looking Time Test trial as the repeated factor. Results indicated a non-significant interaction ($F(2, 51) = 1.13$, $p = .337$, partial $\eta^2 = .042$); however, significant main effects of diagnosis ($F(1, 51) = 5.44$, $p = .007$, partial $\eta^2 = .176$) and trial emerged ($F(1, 51) = 8.22$, $p = .006$, partial $\eta^2 = .139$) indicating that overall, toddlers looked longer towards the first Habituation trial compared to the first Looking Time Test trial. Gabriel's procedure post hoc test (used due to differences in sample size) indicated a significant difference between the DD and TD group such that the difference in looking time towards the first Habituation trial vs. the first Looking Time Test trial was greater in the TD group ($p = .006$). The ASD group demonstrated an intermediate amount of difference in looking time (ps>.05).

Table 8. *Looking Time Characteristics for TAT Looking Time Test Trials by Trial Number and First Hinderer and Helper Trial for the Final Sample*

TAT Looking Time Test Trial	Total (*n* = 54)	ASD (*n* = 15)	DD (*n* = 19)	TD (*n* = 20)
Trial 1				
Mean	8.99	9.83	6.99	10.25
Range	0.10-30.00	0.10-29.40	0.10-19.00	1.20-30.00
Standard Deviation	7.19	8.13	4.85	8.19
Trial 2				
Mean	5.82	6.58	5.36	5.69
Range	0.10-20.90	0.10-13.10	1.10-17.60	0.10-20.90
Standard Deviation	4.47	4.28	4.03	5.12
Trial 3				
Mean	3.64	4.85	2.76	3.56
Range	0.10-18.00	0.10-18.00	0.10-6.90	0.10-11.10
Standard Deviation	3.49	4.97	1.96	3.12
Trial 4				
Mean	4.99	4.26	5.41	5.30
Range	0.10-21.60	0.40-13.80	.10-20.10	0.10-21.60
Standard Deviation	5.28	4.08	4.08	6.09
Trial 5				
Mean	3.53	4.04	4.35	2.63
Range	0.10-14.40	0.60-13.20	0.10-14.40	0.10-7.90
Standard Deviation	3.66	4.06	4.06	2.43
Trial 6				
Mean	4.40	3.65	5.23	5.04
Range	0.10-22.80	0.20-8.30	0.10-22.00	0.10-22.80
Standard Deviation	5.15	2.96	2.96	6.39
First Hinder Trial				
Mean	7.38	9.76	4.13	7.30
Range	0.10-29.60	0.10-29.40	1.10-17.60	0.10-29.60
Standard Deviation	6.46	7.52	7.52	7.13
First Helper Trial				
Mean	7.42	6.65	4.84	8.63
Range	0.10-30.00	0.10-19.50	0.10-19.00	0.50-30.00
Standard Deviation	5.92	5.32	5.32	7.23
First Habituation Trial				
Mean	13.04	13.83	8.57	7.56
Range	0.10-30.00	5.00-30.00	.10-29.60	3.30-30.00
Standard Deviation	8.43	7.06	7.56	8.53

3.3.11 *Aim 3b.*

To measure whether looking time predicts a diagnosis of ASD, a logistic regression was conducted with difference in looking time towards the hinderer and helper during Looking Time Test Trials as the predictor, and ASD status as the outcome. Difference in looking time did not significantly predict ASD status, $b = .108$ Wald $\chi^2(1) = 1.57$, $p = .210$.

3.3.12 *Aim 3c.*

To measure whether looking time predicts ASD symptom severity, a linear regression was conducted with difference in looking time towards the hinderer and helper during Looking Time Test Trials as the predictor, and ADOS-2 Severity Score as the outcome. This relationship was not significant ($F(1,31) = .169$, $p = .684$, $R^2 = .005$, $b = .064$).

3.3.13 *Aim 3d.*

To measure whether there are group differences in shape choice on the Choice Test, a chi-square analysis was conducted with the 3x2 matrix of diagnostic group and shape choice; however, this analysis was not significant ($X^2(2, N = 51) = 1.883$, $p = .390$). Results are depicted in Figure 7.

3.3.14 *Aim 3e.*

Additionally, to explore whether shape choice predicts an ASD diagnosis, a likelihood ratio was computed with shape choice as the predictor and ASD status as the outcome; however, this was not significant, LR(1, $N = 51$) = .667, $p = .411$.

3.3.15 Aim 3f.

To examine whether ASD symptom severity differs based on shape choice, a *t*-test was conducted with shape choice as the independent variable and ADOS-2 CSS as the dependent variable; however, there was no significant effect, *t*(26) = .974, *p* = .339.

3.4 Aim 4: Exploratory analyses

3.4.1 Aim 4a.

To explore whether performance differs by task modality and whether this differs by diagnostic group on the Looking Time Test*,* a 3x2 mixed factorial ANOVA was conducted with diagnostic group as the between factor, mean difference in looking time towards the Hindering and Helping test trials during the Looking Time Test on the PAT and TAT as the repeated factor, and age at Session 1 included as a covariate. The main effect of diagnosis (*F*(2, 48) = 0.30, *p* = .739, partial η^2 = .013), and modality *(F*(1, 48) = 0.609, *p* = .439, partial η^2 = .013) were not significant. The interaction between the effect of diagnosis and trial was also not significant *(F*(2, 48) = 0.760, *p* = .388, partial η^2 = .016). Results are depicted in Figure 8.

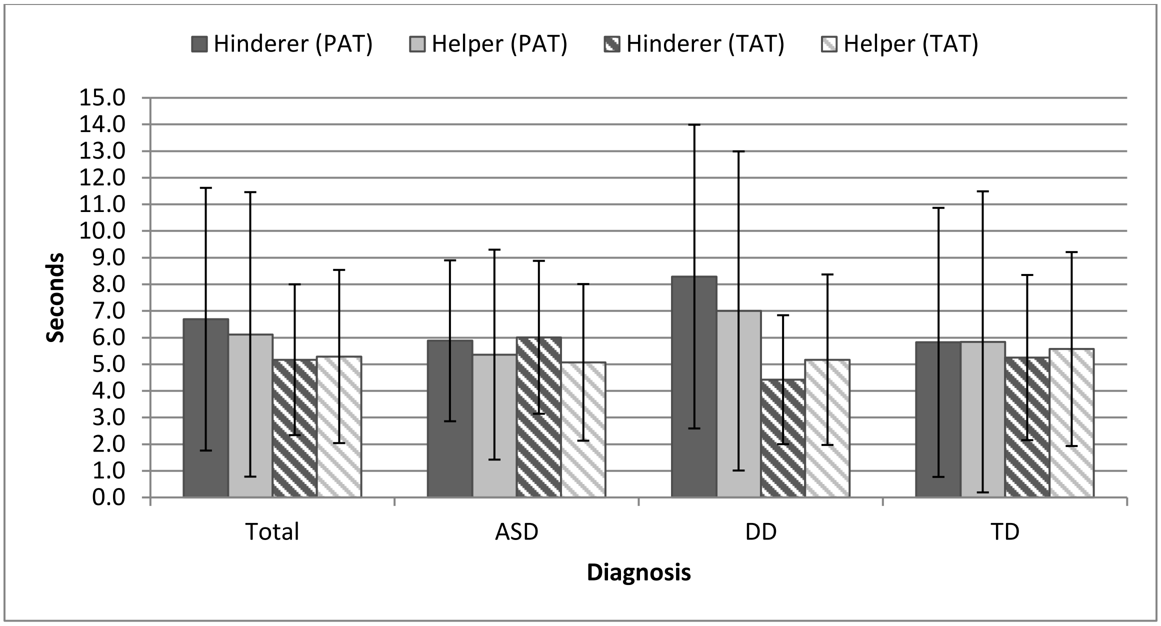

Figure 8. Mean looking time towards Helping and Hindering trials during the PAT and TAT Looking Time Test.

3.4.2 Aim 4b.

To explore whether modality influences task performance and whether this differs by diagnostic group on the Choice Test, a 3x2x2 loglinear analysis was conducted including diagnostic group and shape choice for the PAT and TAT. The three-way loglinear analysis produced a final model that retained the main effect of TAT ($X^2(10) = 12.132$, $p = .276$). Follow up chi-square indicated that toddlers were more likely to choose the hinderer than the helper across diagnostic group during the TAT ($X^2(1) = 8.65$, $p = .003$, odds ratio = 2.4). Results depicted in Figure 7.

4 DISCUSSION

There is a pressing need to examine early markers of ASD in order to improve early detection and diagnosis. This study addressed this need by examining social attribution in toddlers at risk for autism and typically developing toddlers. This study adapted a habituation-based violation of expectation paradigm, designed to study social cognitive development in infants, to study social attribution and social evaluation in toddlers with and without autism in an effort to understand the utility of social attribution as an early behavioral marker in ASD. This study also sought to examine the feasibility of a novel touchscreen adaptation of this measure to inform the use of touchscreen technology to study cognition in toddlers with and without delays and in clinical tool development. Furthermore, this study examined habituation characteristics to better understand performance on subsequent measures of social attribution, potential habituation-based differences across diagnostic groups, and inform the use of this paradigm in toddler populations.

4.1 Habituation

With regard to habituation, it was hypothesized that the proportion of toddlers to reach habituation criterion during the Habituation Phase would be similar across diagnostic groups and modality. These predictions were supported by the habituation data during the learning phases of both PAT and TAT tasks. It was also predicted that children with ASD would require increased exposures to reach habituation criterion compared to the TD and DD groups; however, this was not supported given that all toddlers reached habituation criterion after a similar number of trials. It was also predicted that children with ASD would demonstrate reduced engagement (i.e., less average looking time per habituation trial) compared to the TD and DD groups for the PAT, but would demonstrate increased engagement on the TAT compared to the TD and DD groups. This hypothesis was not supported. Specifically, typically developing toddlers demonstrated longer mean looking time per habituation trial compared to the DD group only. As such, toddlers with DD demonstrated reduced engagement compared to typically developing toddlers on the PAT and TAT, but individuals with ASD did not demonstrate this pattern.

First, it is important to address the finding that the proportion of toddlers who habituated was similar across diagnostic groups and tasks. This suggests that differences by diagnostic group observed on outcomes (i.e., social evaluation measures) were likely not due to differences in whether children reached habituation criterion or not. Given that habituation across diagnostic groups is similar on both puppet and touchscreen versions of this task, it seems reasonable to continue to explore the use of touchscreen technology in future research. This is a valuable finding, given that the materials and space needed for a puppet task are not suitable for clinical settings, whereas a touchscreen can be flexibly used in a wide range of community settings.

Contrary to previous research (Vivanti et al., 2017; Webb et al. 2010), individuals in the ASD group required a similar number of presentations to habituate to the stimuli compared to the other groups. This is useful because it also supports that outcomes in the current study (i.e., measures of social attribution and social evaluation) are not associated with differences in habituation characteristics such as increased exposures during the learning phase. Notably, there were differences in the habituation paradigms used in the current study and these previous studies, which illustrates that design variables may impact whether toddlers with ASD habituate at similar rates or not to other groups of children. However, given that current findings contradict previous findings, this could also indicate that toddlers with ASD did not habituate to the scene and that reaching habituation criterion was associated with task boredom rather than habituation specifically in the ASD group. In retrospect, it would have been ideal to have some kind of 'habituation check' (e.g., examine looking time towards an out-of-category stimulus such as a novel shape) after habituation criterion was reached and prior to showing the violation of expectation task to ensure toddlers reached habituation criterion because learning occurred in order to rule out alternate explanations.

Interestingly, the typically developing group tended to look significantly longer than toddlers in the DD group during the Habituation Phase in both modalities, whereas toddlers in the ASD group looked an intermediate amount. Notably, when this analyses was conducted excluding the six children with MSEL VR <12 months, this main effect was no longer significant. Although this drop in significance could be due to a loss of power, it should also be considered that developmentally delayed toddlers with lower mental age looked at the scenes on average for a shorter duration than developmentally delayed toddlers with a higher mental age. However, mean looking time per habituation trial for toddlers in the DD group with MSEL VR <12 months vs. ≥12 months was not significantly different for the PAT or TAT, so this is most likely a power issue.

The finding that toddlers with non-ASD delays were less engaged during habituation than typically developing toddlers is in contrast to expectation given previous research examining habituation characteristics in typically developing and developmentally delayed toddlers. Specifically, previous research has shown that habituation characteristics are similar between children with typical development and those with non-ASD developmental delays (Vivanti et al., 2017; Webb et al., 2010). Furthermore, some research has shown reduced engagement during habituation for individuals with ASD compared to typically developing and developmentally delayed toddlers (Vivanti et al. 2017); however other research suggests that toddlers may demonstrate increased learning and engagement through computerized methods (Heimann, 1995; Moore and Clavert, 2000; Shane et al., 2008). In the current study sample, parents reported significantly more screen media use in the ASD sample vs. the DD and TD samples, which also supported our initial hypothesis; however, in contrast to both of these bodies of literature and the current findings, we found no difference in engagement, as measured by increased looking time per habituation trial, across tasks in the ASD group.

This finding also adds to the small body of literature examining visual habituation in ASD and is only the second study to our knowledge examining visual habituation to social stimuli in toddlers with ASD. Given that this finding contradicts the previous study examining habituation in ASD (Webb et al., 2010), future research is warranted to examine the habituation characteristics in ASD to inform future research design and to understand the potential clinical utility of potential differential habituation in ASD.

4.2 Social attribution

In the current study, social attribution was measured by increased looking towards an unexpected event (i.e., increased looking towards the climber moving towards the hinderer vs. the helper during the Looking Time Test). It was predicted that typically developing toddlers and toddlers with non-ASD delays would demonstrate evidence of social attribution, whereas toddlers with ASD would show reduced social attribution and that evidence of social attribution skills would predict a diagnosis of ASD and symptom severity; however, these hypotheses were not supported. In fact, none of the groups demonstrated increased looking towards the surprising event on the PAT or the TAT. Consistently, performance did not predict clinical variables including a diagnosis of ASD or ASD symptom severity.

The current study did not replicate previous studies examining social attribution in typically developing infants within this typically developing toddler sample. Furthermore, findings did not replicate pilot findings examining this task in typically developing toddlers (Gonsiorowski et al., 2015).

4.2.1 Why were previous findings from TD samples not replicated?

It is important to think about why current findings do not replicate previous studies in typically developing populations as null findings in the typically developing sample make it difficult to interpret performance in the ASD and DD samples. Specifically, it is important to note that just because toddlers did not demonstrate a difference in mean looking time toward the expected and unexpected event during the Looking Time Test, does not mean that toddlers did not engage in social attribution during the Habituation Phase. Thus, it is important to consider potential limitations of the social attribution measure as well as the possibility that social attribution was not elicited during habituation.

Previous studies examining visual habituation suggest that an absence of an immediate novelty preference does not necessitate that infants failed to learn during habituation (Pascalis & de Hann, 1998). In fact, a number of studies have demonstrated that infants and adults who do not show a novelty preference immediately after habituation may demonstrate a novelty preference at a later time (Bahrick & Pickens, 1995; Courage & Howe, 1998) or demonstrate learning during habituation in other ways (Richmond, Colombo, & Hayne, 2007; Snyder, Blank & Marsolek, 2008). In a study examining visual habituation in toddlers, Webb and colleagues (2010) found that typically developing toddlers did not show a novelty preference after visual habituation. The authors wondered whether reaching habituation criterion evidenced boredom with the task rather than learning; however in examining peak looking towards a subsequent novel out-of-category stimulus, it was discovered that toddlers looked for a similar amount of time to the initial stimuli. This provided evidence that learning occurred during habituation, but that the measurement (i.e., within-category novelty preference) was not an adequate measure of learning in this toddler group. In the current study, there is evidence to suggest that the Looking Time Test may not be a valid measure of social attribution due to boredom. Specifically, toddlers showed decreased looking towards the first Looking Time Test trial compared to the first Habituation trial in both the PAT and the TAT, which shows that they were increasingly uninterested in attending to the task even though a novel stimulus was presented (i.e., a new scene). These data are further supported by the fact that there were significant differences in looking time towards the first and second Looking Time Test trial presented, regardless of trial type (i.e., Helper or Hinderer), and that toddlers decreased looking time toward test trials in a linear fashion. Our data suggest that if social attribution did indeed occur in the TD group during Habituation Phase, boredom may have been a more powerful force than the impact of social attribution on differential looking patterns during the Looking Time Test. In retrospect, it would have been ideal to have another measure of social attribution later on, as some research suggests. In fact, several previous studies (i.e., Hamlin et al., 2007; Kuhlmeier et al., 2003) counterbalanced task order (i.e., Looking Time Test and Choice Test); however the social attribution measure always preceded the choice measure in the current experiment. Previous studies did not report on whether there were differences in social attribution performance based on task order, but future research should examine the potential impact of time post habituation on the measurement of social attribution in toddlers and should explore other means of social attribution measurement.

It is also important to draw attention to the mixed findings in previous studies utilizing this paradigm. Kuhlmeier and colleagues (2003) reported that infants preferred to look at the “coherent continuation” of social scene in a novel context (i.e., the Looking Time Test) as evidenced by increased looking towards the climber moving towards the helper vs. the hinderer. In contrast, Hamlin and colleagues (2007, 2010) found that infants looked longer towards the surprising event (i.e., when the climber moved towards the hinderer vs. the helper). Hamlin and colleagues claimed to have replicated previous findings by Kuhlmeier and colleagues, yet did not address that these findings are in fact opposite. Hamlin and colleagues’ findings are more consistent with a traditional violation of expectation paradigm (i.e., infants look longer toward an unexpected event); however, Kuhlmeier and colleagues argued that the paradigm is actually not a classic violation of expectation paradigm experiment given that the test phase occurs in a completely novel scenario. Kuhlmeier cited that it is not always understood in what conditions infants will look longer towards a coherent or incoherent event and therefore, while they predicted there would be a difference in looking time, they did not make a priori predictions about directionality. In examining the methodology between these three studies and the current study, there are some differences related to presentation modality (i.e., computerized vs. live puppet), presence of googly eyes to scaffold the social interpretation of the shapes, and Looking Time Test context (i.e., Kuhlmeier used an open plain in which shapes did not follow laws of physics vs. Hamlin used an grassy knoll in which shapes followed laws of physics) that may have contributed to this inconsistency in the directionality of findings (Table 9). These differences in the stimuli could explain differences in the direction of the looking time differences across these studies; however this does not explain why no difference emerged across task modalities in the current study. Although predictions made in the current study were based on findings consistent from Hamlin and Gonsiorowski, it is possible that there were some children who looked more towards the surprising event and some who looked more towards the coherent event and that this tendency is child-specific, such as perseveration, or rigidity. For example, we may expect some individuals with ASD to look longer towards the coherent continuation vs. the unexpected event given increased perseveration and rigidity in this clinical sample (Landry & Al-Taie, 2016). Notably, we did not find that difference in looking time either way predict a diagnosis or symptom severity; however this null finding may be due to the high heterogeneity of the ASD population. Nonetheless, our findings do not support child-specific variables impacted any group given that there was a significant decline in looking time between the first and second test trial, regardless of test trial order. Future research should examine whether this tendency is child-specific and perhaps based on other variables such as perseveration.

Table 9. *Methodological Similarities and Differences in Previous Studies Examining Social Attribution and the Current Study*

Study	Age	Modality	Googly Eyes?	Looking Time Test Context	Looking Time Test Performance
Kuhlmeier et al., 2003	Infants (12 ms)	Computer	No	Open field (climber floated)	Helping > Hindering (preference)
Hamlin et al., 2007	Infants (10 ms)	Puppet	Yes	Grassy knoll	Helping < Hindering (surprise)
Gonsiorowski, et al., 2015	Toddlers	Puppet	Yes	Open field (climber floated)	Helping < Hindering scenes (surprise)
Current Study	Toddlers	Puppet & computer (touchscreen)	No	Grassy plain	Helping = Hindering (both modalities)

Note. Ms = months

Another possibility is that typically developing toddlers engaged in social attribution and demonstrated gaze differences during the Looking Time Test that were not captured by the utilization of gross looking time. For example, Fawcett and Liszowski (2012) utilized anticipatory gaze as an outcome instead of overall mean looking time differences in a typically developing infant sample and found that infants made significantly more anticipatory gazes toward the helping vs. hindering character. Notably, anticipatory gaze is frequently used in studies with children beyond infancy and perhaps would be a better method of capturing social attribution in toddler samples. However, in considering clinical applications, current eye tracking methods are expensive and some data suggest that they may contribute to increased missing data in minimally verbal individuals with ASD compared to other technologies (Plesa Skwerer et al., 2016).

Furthermore, it is important to recognize that Gonsiorowski and colleagues (2015) did find differential looking time across test trial types in the Looking Time Test in a toddler sample, but that their Looking Time Test in that study most closely represented the Looking Time Test from the Kuhlmeier and colleagues study (2003), as the scene included an open field with the climber floating upwards towards the helper or hinderer. Given that toddlers in that study were habituated to a scene comprised of a green hill, then tested in a context that did not include any green ground at all, it may be that the change to the scene context somehow scaffolded a differential looking time response. In other words, perhaps the test scene context in the current study was too similar to the habituated context (i.e., both scenes had a green ground and featured the same shapes) such that a differential looking time response indicating social attribution occurred was not elicited. Webb and colleagues (2010) finding that toddlers in their sample did not demonstrate learning during habituation when the test was too similar to the habituation context (i.e., within category), but did when the stimuli were different enough (i.e., out of context) support this idea.

Although the Looking Time Test may not have captured social attribution, we also cannot rule out the fact that social attribution may not have occurred during learning in this toddler sample. First, the primary difference between the current study and previous studies using social attribution/evaluation paradigms in typically developing children is that the majority of previous studies were conducted in infants (i.e., children 12 months and younger). In fact, it is important to note that habituation paradigms in general are utilized to study infant development, whereas studies examining toddler development often require more of an active role of the child (Keen, 2003). Besides data from Gonsiorowski and colleagues (2015), there are only two other known study utilizing a habituation paradigm in toddlers (Vivanti et al., 2017; Webb et al., 2010). Given the differences between infants and toddlers (e.g., higher developed motor, communication, and cognitive abilities), there are likely differences in the way that habituation paradigms function in toddlers. As mentioned, we had no way to know whether toddlers actually habituated to the scenes vs. whether habituation criterion was reached due to other factors such as boredom. As mentioned in the discussion of the habituation results, it would have been ideal to have a 'habituation check' of some kind to understand whether toddlers learned something (perhaps something about the non-social features of the scene), prior to asking them to demonstrate whether they learned something social (i.e., engaged in social attribution).

It is also important to highlight that the methodology used in previous studies are mixed. With regard to the only other study utilizing this paradigm in toddlers with and without ASD (Gonsiorowski, et al., 2015), it is notable that the shapes featured googly eyes, which may have scaffolded social attribution. Although previous studies have shown typically developing infants no longer require googly eyes at 12 months of age in the computerized version of the task (Kuhlmeier et al., 2003), it could be that the potential aforementioned differences that place toddlers at a disadvantage to demonstrate social attribution via this habituation paradigm may have scaffolded toddlers enough to demonstrate social attribution in that study. For example, perhaps toddlers demonstrated increased boredom with the task compared to infants, which reduced the total amount of time they looked at the task, but the presence of googly eyes scaffolded performance such that less time was needed in order to process the scene. Unfortunately, this hypothesis cannot be examined because previous studies do not report sum of looking time during habituation. Sum of looking time across habituation would be important in understanding this hypothesis since the average number of trials to habituation is dependent on a ratio of looking time across time. Although the mean number of trials required to habituate in the TD group is generally consistent with previous studies in infant samples, the sum of looking time could be very different and previous studies do not report this. Furthermore, mean looking time across trials does not help to answer this question since children required different numbers of trials to habituate. In line with this, it should be mentioned that six toddlers with a mental age of less than 12 months were included in the final sample in the current study; however, this likely did not contribute to null findings given that when these children were excluded, there were not changes in significance or effect sizes of results related to social attribution.

Furthermore, the mean number of habituation trials required for the PAT and TAT were comparable to previous studies in infants (Hamlin et al., 2007; Kuhlmeier et al., 2003); however, 22.8% and 31.6% of toddlers in the current study did not meet habituation criterion within 12 trials and it is questionable whether these children truly learned everything they could have from the social scene. While previous studies did not exclude children who did not meet habituation criterion, they did not report how many children included in analyses did not meet habituation criterion. If the rate of toddlers who did not meet the habituation criterion was higher in the current study compared to previous research, this may also contribute to differences in study findings.

Notably, there were no substantial differences in fuss out rates or compliance rates on the Looking Time Test and Choice Test across diagnostic groups; however, anecdotally, there was considerable variability in the behavior of children within groups, including the TD group, during the Habituation and Looking Time Test Phases. Specifically, some children looked for long periods of time before looking away whereas other children looked away then back to the display frequently within one trial. Notably, these differences are not represented in the average total looking time within each Habituation or Looking Time Test trial based on the end trial criterion. Although these data were not pertinent to the current study, future research may examine the number of looks away from the scene during Habituation and Looking Time Test Phases to examine differences that may exist across diagnostic group or modality, or even differences within groups based on other individual characteristics (e.g., attention, NVIQ), as these differences may impact the ability to measure social attribution using this paradigm in toddlers. These characteristics also may serve useful as behavioral markers of diagnoses or other clinical characteristics like attention.

Furthermore, special attention was paid to ensure that the helping and hindering conditions had minimal differences between them, such as the number of touches between shapes, and the time shapes touched one another, for both the PAT and the TAT. There is potential that the tight control of these conditions – to reduce potential confounds – may have minimized the naturalistic motion cues characterizing both scenes that may have been present in previous studies, especially in the TAT, where these variables were most tightly controlled given the ability to program these quantifiers into the stimuli.

4.2.2 What can we say about social attribution in ASD?

Null findings on the Looking Time Test make it difficult to interpret the lack of differential looking time across test trial type (i.e., helping vs. hindering) in the current study.

Again, it is possible that typically developing toddlers did engage in social attribution, but due to potential measurement limitations of the Looking Time Test, toddlers with ASD and non-ASD delays may also have engaged in social attribution. Notably, preliminary findings from Gonsiorowski and colleagues (2015) demonstrated that toddlers with ASD evidenced social attribution during the Looking Time Test, which supports that toddlers in the current study with ASD, also may have engaged in social attribution, but that subtle differences in measurement methods may have contributed to this contrary differential finding.

Nonetheless, findings from the current study do not support differential social attribution across toddlers with ASD, non-ASD delays, and typical development. Future research is needed to tease apart the potential differences in findings in order to best understand why this may have happened to inform future research. Additionally, the fact that Gonsiorowski and colleagues (2015) found that toddlers with ASD demonstrated evidence of social attribution, this does not necessarily discount deficits in social attribution as an early behavioral marker in ASD. Specifically, future research should examine the developmental trajectory of social attribution skills utilizing their paradigm in high-risk infants to examine whether infants later diagnosed with ASD demonstrate reduced social attribution compared to typically developing individuals early in life. If so, this would contribute to the literature that some aspects of social attribution are not necessarily deficient in ASD, but delayed. Furthermore, examining this mentalizing task in high-risk infants would serve the need posed by a Chevallier and colleagues (2012) to develop mentalizing tasks appropriate for use in infants to contribute to the research examining social motivation vs. social cognition theories of ASD.

4.3 Social evaluation

Social evaluation was measured by toddlers choosing the helper shape vs. the hindering shape when given an opportunity to play with the shape (PAT) or watch the shape dance on the computer screen (TAT). It was predicted that typically developing toddlers and toddlers with developmental delays would demonstrate evidence of social evaluation; however hypotheses were only partially supported. Specifically, typically developing toddlers showed a preference for the helper shape more often than the hindering shape during the PAT, but they did not show this preference on the TAT. Furthermore, toddlers with ASD demonstrated reduced evidence of social evaluation on PAT compared to the TD group; however, the DD group also did not show a shape preference on this task. No shape preferences were observed on the TAT across diagnostic groups. Shape preference did not predict clinical variables including a diagnosis of ASD or ASD symptom severity in either modality.

The finding that TD toddlers chose the helper more often than the hinderer replicates previous studies examining social evaluation in typically developing infants (Hamlin et al., 2007; Hamlin et al., 2010). Furthermore, this study replicates findings from the only other known study examining this paradigm in an ASD sample (Gonsiorowski et al., 2015) and extended that study with the inclusion of a non-ASD developmentally delayed sample. Notably, Gonsiorowski and colleagues suggested that although both typically developing children and those with ASD demonstrated social attribution, the children with ASD were unable to integrate this information in making a social judgment about the shapes. Although the current study found no difference in looking time towards helping or hindering scenes even for typically developing children, typically developing toddlers demonstrated evidence of social evaluation (i.e., chose helper at a rate greater than chance i.e., 73%). Given that previous studies suggest that social evaluation comes online earlier in development (i.e., three months) compared to social attribution (i.e., 10 months), this taken together with the current finding suggests that social evaluation and social attribution likely demand differential cognitive abilities. In other words, it may not be the case that toddlers with ASD in Gonsiorowski's study were limited in integrating mental state attribution into their social evaluation, but rather social evaluation is not dependent on social attribution (as is also supported by the developmental time course of social evaluation and social attribution).

Given that the mental age of all participants was well above three months, all toddlers should have succeeded in this task, however, social evaluation was reduced in both the ASD group and the DD group. This finding suggests that reduced social evaluation is not specific to ASD. However, we must remember that individuals with developmental delays looked significantly less than typically developing peers during learning (i.e. habituation). Therefore, it could be that developmentally delayed toddlers did not learn from the scene in the same way that typically developing individuals did. Thus, we cannot be certain that the deficit in social evaluation on the PAT is not specific to ASD, since reduced looking during habituation may have impacted the developmentally delayed toddlers performance. This is also supported by previous research in toddlers with developmental delays and typically developing toddlers, which suggests that these groups of children typically do not demonstrate differential habituation characteristics. Therefore, there may be something about this paradigm that differentially dampens learning in the DD group compared to the TD group.

Notably, this is the first study to examine social evaluation in toddlers with non-ASD developmental delays. Results do not support that social evaluation is necessarily reduced in this group given that differential learning may have impacted performance. Considering that findings show that social evaluation was reduced in ASD compared to the typically developing group, despite similar learning, on the PAT, this suggests that social evaluation may be a useful early marker of ASD in toddlers. Future research is needed to understand whether there are differences in social evaluation between the ASD and DD groups in order to inform the utility of reduced social evaluation in differentiating ASD from non-ASD delays.

4.3.1.1 Why was preference for the helper shape not observed in the TD groups during the TAT Choice Test?

Despite the fact that typically developing toddlers looked longer towards each habituation scene compared to the DD group in both task modalities, evidence of social evaluation was only evident in the TD group on the PAT. Thus, if differential learning during habituation between the TD and DD groups impeded performance for the DD group on the PAT Choice Test, this should have also been observed on the touchscreen version as well, especially given minimal performance differences observed across modalities in the current study. It is important to consider whether social evaluation occurred during the TAT Habituation Phase, but was not captured by the TAT Choice Test for the TD group, or whether social evaluation did not occur within the TD group at all on the TAT. Furthermore, contrary to hypotheses and previous research, toddlers evidenced a hinderer bias during the Choice Task when all diagnostic groups were collapsed.

First it is important to highlight the explicit differences between the PAT Choice Test and the TAT Choice Test, which may have led to differential performance on the TAT vs. the PAT if social evaluation occurred during habituation. While the TAT Choice Test was designed to be a computerized version of the PAT Choice Test, it was impossible to equate the child's experience with the shapes across tasks given that the PAT features tangible 3-D objects that the child gets to play with after making a choice, whereas the TAT features 2-D computerized images. In an attempt to provide the child with a similar experience after making a choice, toddlers were shown that after touching a shape on the screen, they would be able to watch that shape dance around on the screen (i.e., Choice Training). Findings from the current study suggest that most children understood the contingency that touching the shape makes the shape dance, as 73% did not require hand-over-hand assistance after the model was provided during Choice Training. These data are supported by studies showing that 15-month old children easily imitate another person touching a button on a touchscreen that makes a sound (Zack, Barr, Gerhardstein, Dickenson, & Meltzoff, 2009). Despite this, watching a shape dance is clearly not the same as having the opportunity to play with the shape. Therefore, this may have contributed to fundamental differences across task modality that may have impacted performance on the TAT regardless of whether social evaluation occurred during habituation.

Other differences between the PAT and TAT Choice Task include that during the TAT, touchscreen training preceded making the choice; however there was no task between the PAT Looking Time Test and toddler's making a choice. This additional task provided the child with additional stimulation and involved learning, which was absent in the PAT and may have dampened the effect of social evaluation on the TAT choice. Furthermore, anecdotally, most toddlers became very excited when they realized the computer screen was a touchscreen. This extreme novelty and excitement appeared increased from the excitement toddlers demonstrated when they were provided the opportunity to choose one of the shapes to play with in the TAT. Of course, this is merely behavioral observation and these differences were not quantified, but it may be informative to systematically observe participants' level of excitement in the future to examine whether this may have dampened the impact of social evaluation on the TAT performance.

It is also important to note that it was observed that many children in the PAT initially chose both shapes with both hands (and were prompted to choose one instead) or after choosing a shape, then reached out for the other. While the frequency of this was not recorded in the current study, it was observed that this appeared to occur less often in the TAT. That is, during the TAT, toddlers typically chose only one shape (i.e., put one hand on or pointed to only one shape) rather than using both hands to touch both shapes. Notably, choosing both shapes after an initial choice was made was not possible in the TAT since the other shape disappeared as soon as the initial shape was touched. Again, no data were collected on these observations; however future research should examine the frequency of toddlers choosing both shapes in the PAT as this contradicts Hamlin and colleagues findings that infants as young as three months of age actively avoid the hinderer. Furthermore, it should also be examined whether toddlers indeed choose both shapes less frequently in the touchscreen version and why this may have occurred. One postulation is that toddlers' experience with touchscreens provides them with the contingency knowledge that touching one object results in one outcome. These ideas certainly warrant future research in order to understand potential modality related differences that may impact toddler performance in research paradigms. Furthermore, this informal observation about differential toddler behavior in approaching the Choice Task on the TAT, provided additional support that modality specific differences contributed to differential performance on the PAT vs. TAT Choice Tasks.

Beyond differences across social evaluation measure modality, there is also question as to whether experience with screen media may have impacted either learning during habituation and or performance on the social evaluation measure during the TAT. However, given that hand over hand assistance and performance on the Choice Test was unrelated to daily screen media exposure or daily smartphone and tablet exposure, reduced experience with screen media could not explain current performance deficits in the TD group.

Another piece to consider is whether social evaluation occurred during habituation. There is a body of literature supporting the notion of a “video learning deficit” (Anderson & Pempek, 2005), which posits that children under three years of age demonstrate a reduced ability to learn from people on television compared to people in real life. Despite this, there are no known studies explicitly examining the difference in toddler learning from screen media vs. puppet show contexts and the potential impact of this on developmental research methods.

Lastly, it was highly surprising that toddlers evidenced a hinderer bias when choices were collapsed over diagnostic group. In fact, toddlers chose the hinderer 2.4 times more often than the helper. This finding puts into question whether some kind of social interpretation occurred that elicited a desire to choose the Hinderer to dance. Given that this finding is contrary to previous research utilizing this paradigm (Hamlin et al, 2007, 2010), future research is warranted to better understand whether this finding persists in a larger sample.

4.4 Limitations

A limitation of this study includes the small sample size. Although the sample size in the DD and TD groups were similar to samples presented in previous studies (Hamlin et al, 2007, 2010, Kuhlmeier, 2003), the ASD group was smaller and may have reduced power. Despite this, effect sizes of non-significant effects were small, which suggests that even with additional children added, findings would likely not change.

As previously mentioned, another limitation is the lack of a 'habituation check.' While similar previous studies did not include a habituation check (Hamlin et al, 2007, 2010, Kuhlmeier, 2003), given previous research suggests reduced habituation in individuals with ASD (i.e., a null finding was predicted for the ASD group), it would have been ideal to include a habituation check to ensure that toddlers habituated to the scenes in an effort to reduce the possibility of alternate habituation-related explanations for null findings.

Given that that toddlers demonstrated significantly reduced looking time after presentation of the first test trial across both the PAT and the TAT Looking Time Tests, which likely washed out any potential impact of social attribution in differential looking time, it would have been a good idea to present children with the static test scene (i.e., grassy plain with shapes stationary) for one trial period order to expose toddlers to the novel scene and potentially reduce the impact of novelty of the first test trial on the violation of expectation measurement.

Furthermore, future research should include additional measures of social attribution/evaluation as well as delayed measures to investigate whether measurement utilizing different methods or at a later time would evidence social attribution after the Habituation Phase.

This study is also limited by gross looking time measures. Use of eye tracking would have been helpful in understanding whether anticipatory gaze is a useful measure of social attribution in toddlers; however eye tracking also has it's limitations. While eye tracking would facilitate understanding of social attribution, it is not a useful technology for a clinically meaningful task that would be high disseminated in community settings.

5 CONCLUSION

This is the first study to examine social attribution and social evaluation in a group of toddlers with ASD, non-ASD developmental delays, and typical development. Findings suggest that measures of social attribution traditionally used in infant studies may not be appropriate for use in toddler populations; however, measures of social evaluation may have utility for a broad age range. While this study also demonstrated many similarities across puppet and computerized versions of social attribution and social evaluation measures, subtle differences emerged that may have impacted performance across modality, particularly on a touchscreen administered social evaluation measure. Future research is warranted to examine potential performance differences associated with puppet vs. computerized paradigms in the study of child development. Findings have implications for future research examining theories of social attribution and informing utility of new technologies in toddler research and clinical tool development.

6 REFERENCES

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1-16.

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist*, *48*(5), 505-522

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception London*, *33*, 717-746.

Bahrick, L. E., & Pickens, J. N. (1995). Infant memory for object motion across a period of three months: Implications for a four-phase attention function. *Journal of experimental child psychology*, *59*(3), 343-371.

Bal, E., Yerys, B. E., Sokoloff, J. L., Celano, M. J., Kenworthy, L., Giedd, J. N., & Wallace, G. L. (2013). Do social attribution skills improve with age in children with high functioning autism spectrum disorders?. *Research in Autism Spectrum Disorders*, *7*(1), 9-16.

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

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813-822.

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review.

Barton, M., Boorstein, H., Dumont-Mathieu, T., Herlihy, L., Fein, D. (2012). Toddler ASD Symptom Interview. Self-published.

Bassili, J. N. (1976). Temporal and spatial contingencies in the perception of social events. *Journal of Personality and Social Psychology*, *33*(6), 680.

Bauminger, N., & Kasari, C. (1999). Brief report: Theory of mind in high-functioning children with autism. *Journal of Autism and Developmental Disorders*, *29*(1), 81-86.

Bauminger, N., & Shulman, C. (2003). The development and maintenance of friendship in high-functioning children with autism maternal perceptions. *Autism*, *7*(1), 81-97.

Bayley, N. (1969). Bayley scales of mental development. *New York: Psychological Corporation*.

Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences*, *1*(04), 557-560.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.*, *58*, 47-73.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*(1), B25-B31.

Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: meta-analysis. *Schizophrenia research*, *109*(1), 1-9.

Boraston, Z., Blakemore, S. J., Chilvers, R., & Skuse, D. (2007). Impaired sadness recognition is linked to social interaction deficit in autism. *Neuropsychologia*, *45*(7), 1501-1510.

Bornstein, M. H., & Sigman, M. D. (1986). Continuity in mental development from infancy. *Child development*, 251-274.

Bowler, D. M. (1992). “Theory of Mind” in Asperger's Syndrome Dermot M. Bowler. *Journal of Child Psychology and Psychiatry*, *33*(5), 877-893.

Bowler, D. M., & Thommen, E. (2000). Attribution of mechanical and social causality to animated displays by children with autism. *Autism*, *4*(2), 147-171.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in cognitive sciences*, *12*(5), 187-192.

Calvert, S. L., Watson, J. A., Brinkley, V., & Penny, J. (1990). Computer presentational features for poor readers' recall of information. *Journal of Educational Computing Research*, *6*(3), 287-298.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human brain mapping*, *30*(8), 2313-2335.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*(8), 1839-1849.

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, *12*(3), 314-325.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in cognitive sciences*, *16*(4), 231-239.

Colombo, J. (1993). *Infant cognition: Predicting later intellectual functioning* (Vol. 5). Sage publications.

Courage, M. L., & Howe, M. L. (1998). The ebb and flow of infant attentional preferences: Evidence for long-term recognition memory in 3-month-olds. *Journal of Experimental Child Psychology*, *70*(1), 26-53.

Cullen, H., Kanai, R., Bahrami, B., & Rees, G. (2013). Individual differences in anthropomorphic attributions and human brain structure. *Social cognitive and affective neuroscience*, nst109.

Dittrich, W. H., & Lea, S. E. (1994). Visual perception of intentional motion. *Perception London*, *23*, 253-253.

Fawcett, C., & Liszkowski, U. (2012). Infants anticipate others' social preferences. *Infant and Child Development*, *21*(3), 239-249.

Fox, R. & McDaniel, C. (1982). The perception of biological motion by human infants. Science 218, 486–487.

Frith, C. D. (2007). The social brain?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 671-678.

Frith, C. D., & Frith, U. (1999). Interacting minds--a biological basis. *Science*, *286*(5445), 1692-1695.

Gao, T., McCarthy, G., & Scholl, B. J. (2010). The Wolfpack Effect Perception of Animacy Irresistibly Influences Interactive Behavior. *Psychological science*, *21*(12), 1845-1853.

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive psychology*, *59*(2), 154-179.

Gao, T., & Scholl, B. J. (2011). Chasing vs. stalking: interrupting the perception of animacy. *Journal of experimental psychology: Human perception and performance*, *37*(3), 669.

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: a comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, *19*(11), 1803-1814.

Gonsiorowski, Hrabic, Ludwig, Williamson, and Robins (2015). Implicit Social Evaluations in Toddlers with Autism Spectrum Disorder. International Meeting for Autism Research International Meeting for Autism Research (IMFAR), Salt Lake City, Utah.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557-559.

Hamlin, J.K., Wynn, K., Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923-929.

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*(1), 30-39.

Heberlein, A. S., & Adolphs, R. (2004). Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7487-7491.

Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243-259.

Heimann, M., Nelson, K. E., Tjus, T., & Gillberg, C. (1995). Increasing reading and communication skills in children with autism through an interactive multimedia computer program. *Journal of autism and developmental disorders*, *25*(5), 459-480.

Hood, B., Cole-Davies, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. Developmental Psychology, 39(1), 61–70.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science*, *9*(2), 131-134.

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 549-559.

Johnson, S., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds. *Developmental Science*, *1*(2), 233-238.

Keen, R. (2003). Representation of objects and events: Why do infants look so smart and toddlers look so dumb?. *Current Directions in Psychological Science*, *12*(3), 79-83.

Khulmeier, V., Wynn, K., Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402-408.

Klein, A. M., Zwickel, J., Prinz, W., & Frith, U. (2009). Animated triangles: An eye tracking investigation. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1189-1197.

Klin, A. (1991). Young autistic children's listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *Journal of autism and developmental disorders*, *21*(1), 29-42.

Klin, A. (2000). Attributional social meaning to ambiguous visual stimuli in a higher-functioning Autism and Asperger Syndrome: The Social Attribution Task*. Journal of Child Psychology and Psychiatry*, *41*(7), 831-846.

Klin, A., & Jones, W. (2006). Attributing social and physical meaning to ambiguous visual displays in individuals with higher-functioning autism spectrum disorders. *Brain and cognition*, *61*(1), 40-53.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry*, *59*(9), 809-816.

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, *459*(7244), 257-261.

Landry, O., & Al-Taie, S. (2016). A meta-analysis of the Wisconsin Card Sort Task in autism. *Journal of autism and developmental disorders*, *46*(4), 1220-1235.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule–2nd edition (ADOS-2). *Los Angeles, CA: Western Psychological Corporation*.

Ludwig, Hect, King, Revill, Moore, Fink, Robins (in prep). Validation Study of a Novel Social Attribution Paradigm: The Dynamic Interacting Shape Clips (DISC).

Ludwig,N.N., Williamson, R., Adamson, L.B., Brosnan, S., Jones, W. & Robins, D.L, (2014). Screen Media Questionnaire. Self-published.

McCall, R. B., & Carriger, M. S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child development*, 57-79.

Mazurek, M. O., & Wenstrup, C. (2013). Television, video game and social media use among children with ASD and typically developing siblings. *Journal of autism and developmental disorders*, *43*(6), 1258-1271.

Montague, D. P., & Walker-Andrews, A. S. (2001). Peekaboo: a new look at infants' perception of emotion expressions. *Developmental psychology*, *37*(6), 826.

Moore, M., & Calvert, S. (2000). Brief report: Vocabulary acquisition for children with autism: Teacher or computer instruction. *Journal of autism and developmental disorders*, *30*(4), 359-362.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child development*, *72*(3), 718-735.

Mullen, E. (1995) *Mullen Scales of Early Learning: AGS edition*. Circle Pines, MN: American Guidance Service.

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature*, *395*(6705), 894-896.

Oakes, L.M., Sperka, D. J., & Cantrell, L. (2015). Habit 2. *Unpublished software. Center for Mind and Brain, University of California, Davis.*

Pascalis, O., De Haan, M., Nelson, C. A., & De Schonen, S. (1998). Long-term recognition memory for faces assessed by visual paired comparison in 3-and 6-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 249.

Peterson, C. C., Slaughter, V. P., & Paynter, J. (2007). Social maturity and theory of mind in typically developing children and those on the autism spectrum. *Journal of Child Psychology and Psychiatry*, *48*(12), 1243-1250.

Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*(1), 53-78.

Pierce, K., Conant, D., Hazin, R., Stoner, R., & Desmond, J. (2011). Preference for Geometric Patterns Early in Life as a Risk Factor for AutismPreference for Geometric Patterns and Autism. *Archives of General Psychiatry*, *68*(1), 101-109.

Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, *20*(5), 591-604.

Premack, D., & Premack, A. J. (1997). Infants attribute value to the goal-directed actions of self-propelled objects. Journal of Cognitive Neuroscience, 9(6), 848–856.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(04), 515-526.

Rias, R. M., & Dehkordi, S. R. Computer Game Approach for Children with Autism Spectrum Disorder: A Pilot Study.

Richmond, J., Colombo, M., & Hayne, H. (2007). Interpreting visual preferences in the visual paired-comparison task. *Journal of Experimental Psychology-Learning Memory and Cognition*, *33*(5), 823-831.

Rimé, B., Boulanger, B., Laubin, P., Richir, M., & Stroobants, K. (1985). The perception of interpersonal emotions originated by patterns of movement. *Motivation and emotion*, *9*(3), 241-260.

Robins , D.L., Fein, D., Barton, M. (2009). *Modified Checklist for Autism in Toddlers, Revised (M-CHAT-R)*. Self-published.

Robins , D.L., Fein, D., Barton, M. (2009). *Modified Checklist for Autism in Toddlers (M-CHAT-R) Follow-Up Interview*. Self-published.

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the Modified Checklist for Autism in Toddlers, Revised With Follow-up (M-CHAT-R/F). *Pediatrics*, *133*(1), 37-45.

Rutherford, M. D., Pennington, B. F., & Rogers, S. J. (2006). The perception of animacy in young children with autism. *Journal of autism and developmental disorders*, *36*(8), 983-992.

Sampath, H., Indurkhya, B., & Sivaswamy, J. (2012). A communication system on smart phones and tablets for non-verbal children with autism. In *Computers Helping People with Special Needs* (pp. 323-330). Springer Berlin Heidelberg.

Santos, N. S., Kuzmanovic, B., David, N., Rotarska-Jagiela, A., Eickhoff, S. B., Shah, J. N., ... & Vogeley, K. (2010). Animated brain: A functional neuroimaging study on animacy experience. *NeuroImage*, *53*(1), 291-302.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, *4*(8), 299-309.

Schlottmann, A., & Ray, E. (2010). Goal attribution to schematic animals: do 6-month-olds perceive biological motion as animate?. *Developmental science*, *13*(1), 1-10.

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., ... & Skudlarski, P. (2003). The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1430), 415-427.

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.

Shamay-Tsoory, S. G. (2008). Recognition of ‘fortune of others’ emotions in Asperger syndrome and high functioning autism. *Journal of Autism and developmental disorders*, *38*(8), 1451-1461.

Shane, H. C., & Albert, P. D. (2008). Electronic screen media for persons with autism spectrum disorders: Results of a survey. *Journal of autism and developmental disorders*, *38*(8), 1499-1508.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, *105*(2), 809-813.

Snyder, K. A., Blank, M. P., & Marsolek, C. J. (2008). What form of memory underlies novelty preferences?. *Psychonomic bulletin & review*, *15*(2), 315-321.

Sparrow, S. S., Balla, D. A., Cicchetti, D. V., Harrison, P. L., & Doll, E. A. (1984). Vineland adaptive behavior scales.

Sprong, M., Schothorst, P., Vos, E., Hox, J., & Van Engeland, H. (2007). Theory of mind in schizophrenia Meta-analysis. *The British Journal of Psychiatry*, *191*(1), 5-13.

Stephenson, J., & Limbrick, L. (2015). A review of the use of touch-screen mobile devices by people with developmental disabilities. *Journal of autism and developmental disorders*, *45*(12), 3777-3791.

Stone, W.L., Coonrod, E.E. & Ousley, O.Y. (2000). Brief Report. Screening Tool for Autism in Two-Year-Olds (STAT): Development and Preliminary Data’, *Journal of Autism and Developmental Disorders* 30 (6): 607–12.

Tager-Flusberg, H., & Sullivan, K. (1994). A second look at second-order belief attribution in autism. *Journal of Autism and Developmental Disorders*, *24*(5), 577-586.

Tager-Flusberg, H., & Sullivan, K. (2000). A componential view of theory of mind: Evidence from Williams syndrome. *Cognition*, *76*(1), 59-90.

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception London*, *29*(8), 943-952.

Vasavi, K. P., & Susmitha, R. (2012). Pattern Identification Training System for Children with Autism. *International Journal of Engineering*, *1*(5).

Vivanti, G., Hocking, D. R., Fanning, P. A., Uljarevic, M., Postorino, V., Mazzone, L., & Dissanayake, C. (2017). Attention to novelty versus repetition: Contrasting habituation profiles in Autism and Williams syndrome. *Developmental Cognitive Neuroscience*.

Webb, S. J., Jones, E. J., Merkle, K., Namkung, J., Toth, K., Greenson, J., ... & Dawson, G. (2010). Toddlers with elevated autism symptoms show slowed habituation to faces. *Child Neuropsychology*, *16*(3), 255-278.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1-34.

Woodward, A. L., & Guajardo, J. J. (2002). Infants’ understanding of the point gesture as an object-directed action. *Cognitive Development*, *17*(1), 1061-1084.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin*, *124*(3), 283.

Zack, E., Barr, R., Gerhardstein, P., Dickerson, K., & Meltzoff, A. N. (2009). Infant imitation from television using novel touch screen technology. *British Journal of Developmental Psychology*, *27*(1), 13-26.

7 APPENDICES

7.1 Appendix A

M-CHAT-R^{TM}

Please answer these questions about your child. Keep in mind how your child usually behaves. If you have seen your child do the behavior a few times, but he or she does not usually do it, then please answer **no**. Please circle **yes** or **no** for every question. Thank you very much.

1. If you point at something across the room, does your child look at it? (**For Example**, if you point at a toy or an animal, does your child look at the toy or animal?)	Yes	No
2. Have you ever wondered if your child might be deaf?	Yes	No
3. Does your child play pretend or make-believe? (**For Example**, pretend to drink from an empty cup, pretend to talk on a phone, or pretend to feed a doll or stuffed animal?)	Yes	No
4. Does your child like climbing on things? (**For Example**, furniture, playground equipment, or stairs)	Yes	No
5. Does your child make unusual finger movements near his or her eyes? (**For Example**, does your child wiggle his or her fingers close to his or her eyes?)	Yes	No
6. Does your child point with one finger to ask for something or to get help? (**For Example**, pointing to a snack or toy that is out of reach)	Yes	No
7. Does your child point with one finger to show you something interesting? (**For Example**, pointing to an airplane in the sky or a big truck in the road)	Yes	No
8. Is your child interested in other children? (**For Example**, does your child watch other children, smile at them, or go to them?)	Yes	No
9. Does your child show you things by bringing them to you or holding them up for you to see – not to get help, but just to share? (**For Example**, showing you a flower, a stuffed animal, or a toy truck)	Yes	No
10. Does your child respond when you call his or her name? (**For Example**, does he or she look up, talk or babble, or stop what he or she is doing when you call his or her name?)	Yes	No
11. When you smile at your child, does he or she smile back at you?	Yes	No
12. Does your child get upset by everyday noises? (**For Example**, does your child scream or cry to noise such as a vacuum cleaner or loud music?)	Yes	No
13. Does your child walk?	Yes	No
14. Does your child look you in the eye when you are talking to him or her, playing with him or her, or dressing him or her?	Yes	No
15. Does your child try to copy what you do? (**For Example**, wave bye-bye, clap, or make a funny noise when you do)	Yes	No
16. If you turn your head to look at something, does your child look around to see what you are looking at?	Yes	No
17. Does your child try to get you to watch him or her? (**For Example**, does your child look at you for praise, or say "look" or "watch me"?)	Yes	No
18. Does your child understand when you tell him or her to do something? (**For Example**, if you don't point, can your child understand "put the book on the chair" or "bring me the blanket"?)	Yes	No
19. If something new happens, does your child look at your face to see how you feel about it? (**For Example**, if he or she hears a strange or funny noise, or sees a new toy, will he or she look at your face?)	Yes	No
20. Does your child like movement activities? (**For Example**, being swung or bounced on your knee)	Yes	No

7.2 Appendix B

Screen Media Questionnaire (SMQ)

This questionnaire asks you questions about your child's use of screen media. Your answers will help us understand what s/he does during our study. Please answer all of our questions.

Part I: Total Screen Media Use

In this section please think about the total time your child spends using screen media on an average weekday (i.e., Monday through Friday) and on an average day of the weekend (i.e., Saturday through Sunday). Screen media may include, but is not limited to, television, tablets, smart phones, computers, and other hand-held devices with a screen.

1. On an **average weekday**, how many minutes does your child use any type of screen media?
__________ minutes

2. On an **average day on the weekend**, how many minutes does your child use any type of screen media? ________ minutes

Part II: Screen Media Use by Type

In this section you will be asked about the different kinds of screen media your child uses including television, tablets, smart phones, computers, and other hand-held devices with a screen. You will be asked about how often s/he uses each kind, and how s/he uses each kind. Then, if you have more information about your child's screen media use, or if you child uses a kind of screen media that we did not ask about, please write it in the space provided at the end of the form.

Television Use

1. Does your child use a television? Circle one: Yes or No

2. If yes, what does your child do with the television? Check all that apply:

____Turn it on
____Select channel of choice on own
____Start shows/movies on own (e.g., insert DVD and have it play, stream it)
____Selects and starts video games on own
____Watch movies (may or may not need help selecting channel or starting movie)
____Watch shows (may or may not need help selecting channel or starting show)
____Play video games (may or may not need help setting up video games)
____Watch others play video games
____ Other:____

3. On an **average weekday**, how many minutes does your child use the television?
__________minutes

4. On an **average day on the weekend**, how many minutes does your child use the television?
________ minutes

5. When your child is using the television, how engaged/absorbed with it does s/he seem? Circle one number from 1-5:

1 2 3 4 5

1 = Not very engaged/absorbed.
S/he is very easily distracted
from the television.

3 = Somewhat engaged/absorbed.
S/he is somewhat easily distracted
from the television.

5 = Very engaged/absorbed.
S/he is not easily distracted
from the television.

Stationary Computer/Laptop Use

1. Does your child use a stationary computer or laptop? Circle one: Yes or No

2. If yes, what kind/brand (e.g., Macbook, Dell)? _______________________________

3. What does your child do on the computer? <u>Check all that apply:</u>

_____Turn it on
_____Access websites on own
_____Use websites (may or may not need help accessing websites)
List websites typically used: ____
_____Access games on own
_____Play games (may or may not need help accessing games)
List games typically played:____
_____Access movies/videos/shows on own
_____Watch movies/videos/shows (may or may not need help accessing them)
_____Video chat
_____Access music on own
_____Listen to music (may or may not need help accessing music on own)
_____Read books
_____Other:____

4. On an **average weekday**, how many minutes does your child play with/watch something on the computer? __________ minutes

5. On an **average day on the weekend**, how many minutes does your child play with/watch something on the computer? ________ minutes

6. When your child is using the computer, how engaged/absorbed with it does s/he seem? Circle one number from 1-5:

1 2 3 4 5

1 = Not very engaged/absorbed. S/he is very easily distracted from the computer.	3 = Somewhat engaged/absorbed. S/he is somewhat easily distracted from the computer.	5 = Very engaged/absorbed. S/he is <u>not</u> easily distracted from the computer.

Tablet Use

1. Does your child use a tablet? Circle one: Yes or No

2. If yes, what kind/brand (e.g., iPad, Galaxy, Nexus, Surface)?_____________________

3. What does your child do on the tablet? Check all that apply:

_____Turn it on
_____Access applications on own
_____Use applications (may or may not need help accessing applications)
    List applications typically used:__________________________________
_____Access websites on own
_____Use websites (may or may not need help using websites)
    List websites typically used: ____
_____Access games on own
_____Play games (may or may not need help accessing games)
    List games typically played:____
_____Access movies/videos/shows on own
_____Watch movies/videos/shows (may or may not need help accessing them)
_____Video chat
_____Access music on own
_____Listen to music (may or may not need help accessing music on own)
_____Read books
_____Other:____

4. On an **average weekday**, how many minutes does your child use the tablet?
__________minutes

5. On an **average day on the weekend**, how many minutes does your child use the tablet?
________ minutes

6. When your child is using the tablet, how engaged/absorbed with it does s/he seem? Circle one number from 1-5:

1 2 3 4 5

1 = Not very engaged/absorbed. S/he is very easily distracted from the tablet.	3 = Somewhat engaged/absorbed. S/he is somewhat easily distracted from the tablet.	5 = Very engaged/absorbed. S/he is not easily distracted from the tablet.

Smart Phone Use

1. Does your child use a smart phone? Circle one: Yes or No

2. If yes, what does your child do on the smart phone? Check all that apply:

_____Turn it on
_____Speak to someone (or pretend to speak to someone)
_____Video chat
_____Access applications on own
_____Use applications (may or may not need help accessing applications)
List applications typically used:___________________________________
_____Access websites on own
_____Use websites (may or may not need help accessing websites)
List websites typically used: ____
_____Access games on own
_____Play games (may or may not need help accessing games)
List games typically played:____
_____Access movies/videos/shows on own
_____Watch movies/videos/shows (may or may not need help accessing them)
_____Access music on own
_____Listen to music (may or may not need help accessing music)
_____Read books
_____Other:____

3. On an **average weekday**, how many minutes does your child use the smart phone?
__________ minutes

4. On an **average day on the weekend**, how many minutes does your child use the smart phone? ________ minutes

5. When your child is using the tablet, how engaged/absorbed with it does s/he seem? Circle one number from 1-5:

1 2 3 4 5

1 = Not very engaged/absorbed. S/he is very easily distracted from the smart phone.	3 = Somewhat engaged/absorbed. S/he is somewhat easily distracted from the smart phone .	5 = Very engaged/absorbed. S/he is not easily distracted from the smart phone.

Use of Other Hand-Held Device With a Screen

1. Does your child use any other hand-held devices with a screen (i.e., not a smart phone or tablet)? Circle one: Yes or No

2. If yes, what kind/brand (e.g., PSP, Nintendo 3DS)?___________________________

3. How does your child use the hand-held device (e.g., Can s/he turn it on and access games? How does s/he play with it?) Describe:

4. On an **average weekday**, how many minutes does your child use this hand-held device? __________ minutes

5. On an **average day on the weekend**, how many minutes does your child use this hand-held device? ________ minutes

6. When your child is using the hand-held device, how engaged/absorbed with it does s/he seem? Circle one number from 1-5:

1	2	3	4	5

1 = Not very engaged/absorbed.
S/he is very easily distracted
from the device.

3 = Somewhat engaged/absorbed.
S/he is somewhat easily distracted
from the device.

5 = Very engaged/absorbed.
S/he is <u>not</u> easily distracted
from the device.

If there is any more information that you think is important about your child's screen media use, or if there is another kind of screen media that your child plays with that we did not ask about, please provide that information here (use back if you need more space):