3-6-2017

Temperament in Baby Siblings of Children with Autism Spectrum Disorder

Julia Chen
University of Connecticut - Storrs, julia.chen@uconn.edu

Temperament in Baby Siblings of Children with Autism Spectrum Disorder

Julia Chen

B.A., Columbia University, 2014

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science at the

University of Connecticut

2017

APPROVAL PAGE

Master of Science Thesis

Temperament in Baby Siblings of Children with Autism Spectrum Disorder

Presented by

Julia Chen, B.A.

Major Advisor____

Marianne Barton, Ph.D.

Associate Advisor____

Jeffrey Burke, Ph.D.

Associate Advisor____

Kimberly Cuevas, Ph.D.

University of Connecticut

2017

ACKNOWLEDGEMENTS

Committee

Marianne Barton, Ph.D., Jeffrey Burke, Ph.D., Kimberly Cuevas, Ph.D.

Early Detection Team

Deborah Fein, Ph.D., Thyde Dumont-Mathieu, MD, Harriet Levin, OTL/R,

Lauren Haisley, Dasal Jashar, Kate Bradbury, Emily Moulton, Cara Cordeaux,

Lauren Miller, Yael Dai, Katelynn Porto, Jennifer Donelan

Undergraduate Research Assistants

Participating Children, Families, and Pediatricians

Collaborators at the University of Washington and Vanderbilt University

Funding Support

Eunice Kennedy Shriver National Institute of Child Health and Human Development

(5 RO1 HD039961-05)

TABLE OF CONTENTS

Abstract

Objective: Temperament may be an important early behavioral risk marker for ASD. The current study aimed to investigate temperamental differences and stability of temperament in high-risk baby siblings with and without ASD in toddlerhood and preschool-age; understand the additional benefit of measuring temperament in predicting ASD, beyond measures of ASD symptomology and severity; and understand gender differences in the relationship between temperament and ASD.

Methods: High-risk baby siblings of children with ASD were evaluated at approximate ages 2 and 4. Multiple regression was used to assess the relationship between diagnostic group (ASD, non-ASD), temperament at both time points, and gender. Logistic regression was used to predict diagnostic group membership from temperament, accounting for scores on a measure of ASD symptomology and severity.

Results: Differences between baby siblings with and without ASD were demonstrated on several temperament domains. Poor approach and higher negative mood increased likelihood of ASD diagnostic group membership at age 2, after accounting for ASD symptomology and severity. Most temperament domains were stable from ages 2 and 4, potentially more so for the ASD group. Several gender-specific differences in temperament between the ASD and non-ASD groups were found.

Conclusions: Temperament may be an early behavioral marker of ASD in high-risk baby siblings, and may provide additional benefit for predicting risk for ASD beyond symptom severity. Gender differences in the relationship between ASD and temperament may provide important clues to how behavioral patterns interact with emergence of ASD symptoms differently, or are perceived differently by caregivers, for boys and girls.

Temperament in Baby Siblings of Children with Autism Spectrum Disorder

Autism Spectrum Disorders (ASD) are a group of neurodevelopmental disorders characterized by persistent impairments in reciprocal social interaction and verbal and nonverbal communication, as well as the presence of restricted and repetitive behaviors manifested early in childhood (American Psychiatric Association; APA, 2013). ASD is currently diagnosed based on the presence or absence of specific behavioral symptoms. While ASD can be reliably diagnosed as young as two years of age (Kleinman et al., 2008; Moore & Goodson, 2003), diagnosis of ASD remains a difficult task because of the great variability in onset, course, and manifestation of symptoms, as well as the variability in infant and toddlers' cognitive abilities and adaptive functioning (Bryson et al., 2007; Landa & Garrett-Mayer, 2006). Many researchers have proposed that temperament may be an important construct that provides useful clues about early behavioral patterns related to autism symptoms, and may account for some of the heterogeneity in the presentation of ASD (Garon et al., 2009).

Temperament has been defined in multiple ways (Goldsmith et al, 1987), but theorists generally agree that temperament represents biologically-based behavioral tendencies that are relatively consistent across the lifespan, but that vary in their expression at different developmental stages and in terms of discrete behaviors. One major theory of temperament is Thomas and Chess's (1977) definition of temperament as the "stylistic component of behavior," that is, *how* a child carries out behavior regardless of motivation for behavior or ability level. Thomas and Chess argue that the best measure of temperament is based on a child's response to external stimuli presented within the child's social context. Temperament is thought to emerge in infancy, with aspects of temperament such as positive and negative affectivity and approach behaviors emerging at 2 to 3 months of age, and regulation of attention, emotions and behavior emerging at the end of the first year and developing into toddlerhood (Kochanska, Murraym, & Harlan, 2000; Rothbart, Ahadi, & Evans, 2000).

While temperament is generally considered to be biologically based, it is nonetheless subject to environmental influences, including early relational experiences, parent and family characteristics (including *within* family non-shared environmental influences, such as birth order), and sociocultural context (Cicchetti & Cohen, 1995). Thus, temperament – like other biological and environmental influences – is often viewed as a risk or protective factor for developmental outcomes (Salley, Miller, & Bell, 2012). Indeed, research with typically developing (TD) children has found that early temperament is associated with later cognitive and language ability, social competence, school adjustment, and risk for internalizing and externalizing symptoms (De Pauw, 2010; Salley et al., 2012; Sanson, Hemphill, & Smart, 2004). For example, studies have shown that high social inhibition predicts later internalizing symptoms, and low self-control predicts later externalizing symptoms (Rothbart & Bates, 1998; Shiner et al., 2003). Temperament has also been shown to predict the quality of joint attention behaviors in toddlerhood, an important early skill that builds the foundation for learning, understanding of intentional communication through gaze, perspective taking abilities, and information processing through exploration of the environment (Todd & Dixon, 2010).

Understanding how child temperament interacts with parenting style to predict behavioral trajectories and outcomes can help caregivers adjust parenting behaviors to provide the “best fit” between a child’s temperament and their environment (Paterson & Sanson, 1999, Sanson et al., 2004). For example, children high in frustration, irritability, impulsivity, and low effortful control have been shown to increase, and to suffer the adverse effects of, negative parenting behaviors such high punishment, power assertion, and unresponsiveness (Kiff, Lengua, &

Zalewski, 2011). Alternatively, difficult temperament may elicit greater investment in positive parenting behaviors from caregivers, though these efforts may be difficult to sustain over time (Sanson et al., 2004).

Temperament in ASD

Motivated by this research on the relationship between temperament and emergence of psychopathology, researchers in the past two decades have begun to think about how early temperament plays into the emergence of ASD. While certain temperament characteristics may not be uniquely associated with risk for ASD, temperament may be an additional risk (or protective) factor on top of risk factors such as high genetic risk (e.g. in younger siblings of children with ASD). Some researchers have proposed that temperament can be conceptualized as a "modifier process" (akin to processes such as socialization and cognitive style) that contribute to risk for ASD, and its varied presentations, course, timing of emergence and severity of symptoms, as well as response to treatment (Clifford et al., 2013; Garon, et al., 2015; Mundy, Henderson, Inge, & Coman, 2007). For example, a child who has a tendency toward high negative affect may have more trouble making progress in treatment, and maintaining friendships compared to a peer with similar social deficits (del Rosario, Gillespie-Lynch, Johnson, Sigman, & Hutman, 2014). Likewise, early abnormalities in attention, reactivity, activity level, and behavioral regulation may compromise important prerequisite processes related to the development of social interaction skills found to be significantly delayed by the time of an ASD diagnosis (Zwaigenbaum et al., 2005). Parental perceptions of their child's behavioral difficulties can also influence early parent-child interactions. Some research shows that parents of children with autism who describe their child as "temperamentally difficult" use more physical control strategies to maintain their child's attention (e.g. holding their child in their lap during play; Kasari, Sigman, Mundy, & Yirmiya, 1988).

Recently, temperament has been conceptualized as a potential biologically-based behavioral marker or endophenotype of ASD (Bryson, et al., 2007; Garon et al., 2015). Indeed, temperament is often included in discussions about the Broader Autism Phenotype (Bailey, Palferman, Heavey, & Le Couteur, 1998), or subclinical autism-related symptoms often observed in relatives of individuals with ASD (Clifford et al., 2013). Conceptualization of temperament as an endophenotype has motivated a line of research studying temperament early within the first year of life and prospectively until the time of diagnosis to capture the earliest manifestations of temperament profiles that distinguish children who go on to develop ASD from peers who do not. Along with early motor and sensory atypicalities, certain temperament profiles are thought to contribute to "prodromal symptoms" of ASD that can be identified within the first year (Sacrey, Bennett, & Zwaigenbaum, 2015). These early behavioral signs, while not part of the diagnostic criteria for ASD, can serve as important clues to a child's risk for the disorder.

Research with Baby Siblings of Children with ASD

Baby siblings of children with ASD are at heightened risk of developing autism themselves (10-20% recurrence risk relative to the 1.5% prevalence rate in the general population (Baio, Centers for Disease Control and Prevention, 2014; Szatmari, et al., 2016). Therefore, high risk baby siblings are uniquely valuable for answering questions about early behavioral patterns preceding an ASD diagnosis. Many prospective studies on temperament have utilized baby sibling samples to investigate hypotheses about causal relationships between temperament and autism risk (Shiner & Caspi, 2003). Comparisons of high risk baby siblings with low risk (LR) siblings (children without a family history of ASD) on temperament dimensions can provide information about temperament features related to the genetic risk associated with ASD. Comparing HR siblings who develop ASD with those who do not (hereon HR-ASD and HR-non-ASD siblings, respectively) can further inform about the role of temperament in risk for ASD above and beyond the genetic risk that HR siblings share.

Infancy and Toddlerhood

Previous research comparing HR-ASD, HR-non-ASD, and LR siblings in infancy and toddlerhood shed light on temperament patterns associated with autism. A case series study measuring temperament at 6, 12, and 24 months in nine HR infants who developed ASD suggested that from 6 to 12 months of age, these infants showed greater irritability, intolerance for intrusions, greater negative affect and proneness to distress, as well as significant difficulties being soothed (Bryson et al., 2007). The authors suggested that an early pattern of irritability and distress with inadequate self-regulation may parallel the emergence of autism symptoms. Another study comparing temperament in 54 HR and 50 LR infants prospectively at 7, 14, and 24 months found that HR infants demonstrated reduced surgency (a temperament domain that encompasses positive affectivity and approach behaviors, as well as impulsivity and activity) at 7 and 14 months, and lower effortful control (attentional and behavioral control) at 7, 14, and 24 months (Rothbart, Ahadi, Hershey, & Fisher, 2001). Compared to HR-non-ASD siblings, HR-ASD siblings had increased negative affect at 24 months (Clifford, et al., 2013). Del Rosario and colleagues' (2014) measured temperament in HR siblings at 6, 12, 18, 24, and 36 months with samples at each time point ranging from 7-27 children. While HR-ASD children were found to be more adaptable and approaching than their HR-non-ASD peers at 6 and 12 months, they were less adaptable and approaching by 24 and 36 months, suggesting that certain temperament patterns may not be indicative of risk for ASD until the second year of life.

In another study with 150 HR siblings (65 of whom were followed longitudinally up to age 2), HR-ASD siblings showed marked passivity and decreased activity at 6 months; intense distress reactions, visual fixation on objects, and decreased positive affect at 12 months; and reduced attention shifting, inhibitory control, and positive affect at 24 months compared to HR-non-ASD siblings (Zwaigenbaum et al., 2005). Garon and colleagues (2009) demonstrated in their study of 104 HR siblings that HR-ASD siblings had lower behavioral approach (i.e. higher activity but reduced goal-oriented extraversion, reward anticipation, and attention shifting) compared to HR-non-ASD siblings. In all, this literature lends support to the idea that temperament characterized as early as the first and second years of life can differentiate HR siblings who develop ASD from those who do not. Specifically, these infants and toddlers demonstrate early patterns of decreased positive affect, adaptability, goal-oriented approach behaviors, and attentional and behavioral control, as well as increased negative affect and intensity of distress reactions. Poor regulation of distress likely reflects the contributions of both biologically-based behavioral tendencies, as well as poor attachment to caregivers (that likely stems from the parallel emergence of social interaction difficulties related to ASD, such as atypical early eye gaze patterns in response to social stimuli; Falck-Yterr, Bolte, & Gredeback, 2013) and prevents adequate seeking of caregivers to help regulate distress. Activity levels in HR-ASD siblings seem to fluctuate from the first to second year (Zwaigenbaum et al., 2005; Garon et al., 2009).

Preschool and School-Age

Due to the nature of the goals of previous studies looking at temperament in HR sibling cohorts (i.e. to track temperament patterns prior to an autism diagnosis), few studies have investigated temperament differences between HR-ASD and HR-non-ASD siblings beyond age

2. Only one study has measured temperament in HR siblings up to 36 months (del Rosario et al., 2014), highlighting a gap in our understanding of how temperament differences between HR siblings with and without ASD present at the start of preschool. Temperament during the preschool years may interact with ASD symptoms in important ways that increase or decrease risk for further psychopathology, as well as difficulties in learning, interaction with peers, and exploration of novel environments (Shiner, et al., 2003).

Studies comparing preschool and school-aged children with ASD (not specified as HR siblings) with TD or other-developmentally delayed peers provide some clues about how temperament may differ between children with and without ASD at this age. Studies indicate that children with ASD exhibit lower effortful control and lower positive affect, and are harder to distract from ongoing behavior, less persistent in challenging activities, harder to sooth, and slower to adapt to and approach new people, objects and events compared to TD or developmentally delayed peers (Adamek et al., 2011; Bailey, Hatton, Mesibov, Ament, & Skinner, 2000; Brock et al, 2012; Hepburn & Stone, 2006; Konstantareas & Stewart, 2006; Ostfeld-Etzion, Feldman, Hirschler-Guttenberg, Laor, & Golan, 2015). There have been mixed findings regarding how perceptually sensitive children with ASD are compared to their peers. Some research has found that children with ASD are more sensitive to stimuli (Konstantareas & Stewart, 2006), while others have found that they have reduced perceptual sensitivity and therefore require more stimulation from their environment (Hepburn & Stone, 2006; Ostfeld-Etzion et al., 2015).

Temperament versus ASD symptomology

Most of the literature reviewed thus far has focused on temperament profiles that distinguish children with and without autism. Several gaps in the literature have yet to be addressed. One relates to questions raised in the literature about how conceptually distinct temperament is from ASD symptomology, given potential overlap in temperament and symptom constructs (e.g. poor approach towards novel people). Is a child with ASD displaying a unique temperament profile, or emerging deficits in social motivation? Of course, temperament encompasses a range of behavioral patterns not necessarily incorporated into the diagnostic conceptualization of ASD (e.g. rhythmicity in feeding and sleeping patterns). However, no research has yet investigated how much more information temperament may provide in helping predict risk for ASD, beyond ASD symptomology. If temperament is to be a useful early marker of ASD, understanding how relevant this information is for early detection can facilitate efforts to improve our diagnostic process.

Stability of Temperament

Furthermore, there has been little research on the stability of temperament in children with ASD. We know that in TD children, there is some stability in temperament from toddlerhood to preschool-age: children classified as “exuberant” (high approach/positive affect) at age 2 continue to display high levels of approach behaviors, less shyness, and increased problem behaviors as a preschooler. Temperament may be less stable for children classified as inhibited (low approach/positive affect) or low reactive (moderate approach, low positive and negative affect; Stifter, Putnam, and Jahromi, 2008). Children at the extreme ends of temperament dimensions show most stability in temperament over time (e.g. children who were highly adaptable and persistent tended to stay that way over time; Pedlow, Sanson, Prior, & Oberkliad, 1993; Sanson, Pedlow, Cann, Prior, & Oberklaid, 1996). Thus far, only one study has looked at temperament across time in ASD, and demonstrated that from 6 to 36 months, HR-ASD baby siblings display increasingly poor adaptability, poor approach, higher intensity of emotional reactions, and higher distractibility from ongoing behavior (Del Rosario et al., 2014), suggesting less stability of temperament before age 3.

Gender Differences

Lastly, there has been no substantial effort made to investigate potential gender differences in temperament of children with ASD. Recent discussions among researchers and individuals in the autism community have elucidated concerns about potential under or misdiagnosis of ASD in girls given the under-representation of girls in ASD research (Daniels & Mandell, 2014; Dworzynski, Ronald, Bolton, & Happe, 2012; Gould & Ashton-Smith, 2011). Indeed, in the temperament literature reviewed thus far, gender or sex has usually been co-varied out of analyses, and treated as a potential confound to understanding the relationship between ASD and temperament (e.g. Brock et al., 2012; del Rosario et al., 2014; Garon et al., 2009). This method not only assumes a fundamental difference between boys and girls in all aspects of temperament, but also leads to conclusions based predominantly on male samples. Alternatively, treating gender as a moderator allows us to evaluate how the relationship between ASD and temperament varies as a function of gender.

Specific Aims

The current study has four aims: (1) to characterize differences between HR-ASD and HR-non-ASD siblings in parent-rated temperament at approximate ages 2 and 4; (2) to understand the utility of including temperament as a predictor of an ASD diagnosis at age 2, after accounting for ASD symptomology and severity; (3) to examine the predictability of age 4 temperament from age 2 temperament for HR baby siblings, and whether presence or absence of ASD at age 2 moderates this predictability; and (4) to investigate whether gender moderates the relationship between an ASD diagnosis and temperament at ages 2 and 4. In this study, baby siblings with or without ASD all screened positive on an autism-specific screener. Thus, our comparison group (HR-non-ASD baby siblings) is more similar to the HR-ASD group in terms of risk for autism (including both genetic risk and parental concerns about ASD), compared to most control groups used in previous studies (e.g., TD peers, peers with Down Syndrome; Bailey, et al., 2000; Konstantareas & Stewart, 2006). Comparison of children who share a relatively similar level of genetic risk for ASD, and whose parents have identified autism-relevant concerns, allows us to better isolate temperament for study.

Our hypotheses for the first aim follow findings from previous research on temperamental differences during infancy and toddlerhood between HR baby siblings with and without ASD. We expect that, compared to HR-non-ASD baby siblings, HR-ASD siblings will demonstrate poorer adaptability, less ability to distract from ongoing behavior, less approach towards novel stimuli, and less display of positive affect; as well as higher activity levels, more display of negative mood, and higher intensity of emotional expression. Based on findings about decreased effortful control found in toddlers with ASD, we hypothesize that the HR-ASD group will be less rhythmic and less persistent on challenging activities due to lower behavioral and attentional control. Given mixed findings regarding perceptual sensitivity in toddlers with ASD, we do not have hypotheses about the direction of difference between our HR-ASD and HR-non-ASD groups. We expect to find these differences at both ages 2 and 4.

For our second aim, we hypothesize that temperament domains (found to differ between ASD and non-ASD groups) will predict significant variance in diagnostic group membership at age 2, after accounting for a measure of autism symptomology and severity. We predict that temperament domains with the least overlap with ASD symptomology (e.g. mood, intensity, rhythmicity, and persistence) will demonstrate the most additional utility for predicting diagnostic group membership.

For Aim 3, we hypothesize – as shown in the TD literature – that temperament at age 2 will predict a significant amount of variance in temperament at age 4 for HR baby siblings as a whole in each temperament domain assessed. However, we expect that because more children in the HR-ASD group will score on the extreme ends of temperament scales at both time points, the relationship between age 2 and age 4 temperament will be stronger for the HR-ASD group (as seen in research with TD samples; Pedlow, et al., 1993; Sanson, et al., 1996).

Lastly, we hypothesize that gender will moderate the relationship between diagnostic group and temperament, and will reflect interruptions to parental expectations about gender norms for their child. Studies with TD children have suggested that gender differences may relate to differences in gender role development, and learned emotional expression and regulation patterns (Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006). Thus, we expect that parents will rate their daughter with ASD as more deviant from girls without ASD (i.e. there will be a greater difference between the female HR-ASD and female HR-non-ASD groups, compared to the difference between the male HR-ASD and male HR-non-ASD groups) on temperament domains in which girls are expected to score at a certain extreme (e.g. girls should be highly socially approaching, have low activity levels, and be highly persistent on challenging tasks; Garon, et al., 2015).

Methods

Participants

Data for this study come from a larger study looking at the utility of an autism screening measure – the Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, &

Green, 2001) and its revised version, the M-CHAT-Revised with Follow-up (M-CHAT-R/F; Robins, Fein, & Barton, 2009) for detecting ASD. Data included in the current study were collected between 2005 to 2012, and come from three sites: University of Connecticut (UConn), University of Washington (UWash), and Vanderbilt University (Vanderbilt). Ethical approval of the study was obtained from each university's respective Institutional Review Boards. Children between the ages of 16 and 30 months were recruited through well-child visits at pediatrician sites, Early Intervention sites, or through clinical services and research involving an older sibling with ASD.

The current study looks at a sample of HR baby siblings who received a diagnostic evaluation and had temperament data at either/both Time 1 or Time 2 (see Figure 1 for a diagram summarizing flow of participants). One hundred forty-four participants were included in Time 1 analyses, 73 participants for Time 2 analyses, and 58 participants for analyses addressing Aim 3. Participants were lost to attrition for the following reasons: 1) re-screening measures were distributed to families through mail and participants were asked to mail screeners back, and a majority (69 out of 83) never returned a screener and did not respond to contact attempts; 2) eight screened negative at Time 2, and were therefore not offered a re-evaluation; 3) and six were not offered a rescreening or re-evaluation through the study because the study had ended by the time of their re-screening. (See Results section and Tables 2a to 2c for results of analyses comparing participant characteristics of those included to those excluded from analyses due to missing temperament data or loss to follow up at Time 2).

Diagnosis of the older sibling was confirmed as part of the sibling study through an evaluation comparable to that received by the baby sibling (described below). Baby siblings with a previous ASD diagnosis from an MD or psychologist (and who therefore would not require another diagnostic evaluation through our study), a major medical condition, or a sensory or motor impairment that would prohibit the child from engaging in evaluation tasks, were excluded from the study. Additionally, children were excluded if they were more than three weeks premature, weighed less than 4.4 pounds at birth, if they did not have an older sibling (full or half-sibling) with an ASD, or if their caregiver did not speak either English or Spanish. Characteristics of participants included in different analyses are detailed in the Results section and Tables 3 and 4.

Procedures

Caregivers completed the M-CHAT(-R/F) at their pediatrician's office, at home, or at the research site. Completed screeners were then scored by research staff. Children who passed the screener were screened again at approximately 48 months of age with the M-CHAT(-R/F). Caregivers of children who failed the screener were contacted for a follow-up phone interview to review the questions their child had failed in more detail. Items with missing or ambiguous responses (e.g., more than one response to an item) were also considered failed items for the purposes of qualifying for the follow-up phone interview. If the child continued to screen positive after the phone interview, the family was invited to their respective research site for a free developmental and diagnostic evaluation. A licensed clinical psychologist or developmental pediatrician and a graduate student or trained research assistant from the clinical psychology PhD programs at the respective research sites conducted the evaluations. Families received a free diagnostic evaluation as part of their participation in the study, and were therefore not compensated monetarily.

Time 1 (Age 2) Evaluation. All parents provided written consent for their own and their child's participation prior to the start of the evaluation. Prior to their child's first evaluation, caregivers were mailed and asked to complete several measures at home, including the Toddler Temperament Scale (TTS; Fullard, McDevitt, & Carey, 1984) and history form. At the evaluation, clinicians interviewed caregivers using the Toddler ASD Symptom Interview (TASI; Barton, Boorstein, Dumont-Mathieu, Herlihy, & Fein, 2012) and Vineland Adaptive Behavioral Scales (VABS; Sparrow, Balla, & Cicchetti, 1984 or VABS-II; Sparrow, Cicchetti, & Balla, 2005). Children's cognitive abilities were assessed using the Mullen Scales of Early Learning (MSEL; Mullen, 1995) and their social and communication skills using the Autism Diagnostic Observation Schedules (ADOS; Lord et al., 2000). Based on both parent report and observations during the evaluation, the clinician rated autism-related symptoms and severity on the Childhood Autism Rating Scale (CARS; Schopler, Reichler, DeVellis, & Daly, 1980).

Children were diagnosed based upon scores on autism-specific diagnostic measures (i.e., ADOS, TASI, and CARS), in conjunction with clinical judgment according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; APA, 2000) criteria. Scores on the MSEL and VABS were also used when making non-ASD diagnoses (diagnostic groups are described in more detail below). Immediately following the evaluation, short feedback sessions were conducted to provide families with testing results, including diagnoses and preliminary recommendations. Six to eight weeks following the evaluation, a comprehensive report detailing the results of the evaluation, interpretation of the child's strengths and weaknesses, as well as appropriate recommendations, was mailed to the family.

Time 2 (Age 4) Evaluation. All families who attended evaluations at Time 1 were invited back for a free, follow-up evaluation when their child was approximately four years old. Some participants (i.e. 20 of 73 [27%] – those recruited through Vanderbilt, and UWash during screening with the M-CHAT-R/F) were rescreened and invited for a re-evaluation around 36 months. Prior to attending a re-evaluation for their child, caregivers were again mailed a packet of measures, including the Behavioral Styles Questionnaire (BSQ; McDevitt & Carey, 1996). Evaluation procedures like those at the Time 1 evaluation were carried out. Diagnostic tools appropriate for children aged 4 to 5 were used (i.e. the Autism Diagnostic Interview-Revised, Lord et al., 1994, instead of the TASI).

Diagnostic Groups. To answer our research questions investigating differences in temperament between HR baby siblings with and without a diagnosis of ASD, the Time 1 and Time 2 samples were each broken down into diagnostic groups: The *ASD* group consists of baby siblings who received an ASD diagnosis – that is, Autistic Disorder or Pervasive Developmental Disorder, Not Otherwise Specified (PDD NOS). The *non-ASD* group consists of baby siblings who received one of the following diagnoses or classifications: Developmental Delay, Developmental Language Delay, Other diagnosis (other DSM or ICD diagnoses not specified above), No Diagnosis (demonstrates sub-clinical symptoms), or Typical Development. It is important to note that children classified as Typical Development may not necessarily represent "typically developing" children of this age, as their parents had indicated sufficient concern about their development to render a positive screen on the M-CHAT(-R/F).

Measures

Researchers are continuing to work towards a consensus on the definition of temperament – debating how narrow or broad the construct is, and which domains of behavior it captures (Goldsmith, et al., 1987; Zentner & Bates, 2008). In the autism literature, some research (e.g. Bryson, et al., 2007; Garon, et al., 2009, 2015; Clifford et al., 2013) have used temperament measures based on Rothbart and colleagues' theory of temperament, which focuses on emotional and behavioral reactivity and regulation (e.g. the *Infant Temperament Scale*; Gartstein &

Rothbart, 2003; Rothbart, 1981). Other researchers (e.g. Del Rosario, et al., 2014; Hepburn & Stone, 2006) have used the Carey Temperament Scales (CTS), based theoretically on Thomas and Chess's (1977) theory of temperament. We have also chosen to use the CTS for this study, as it is most comprehensive in capturing the multiple domains of child temperament identified in existing temperament measures (e.g. behavioral inhibition/approach-withdrawal, mood, activity level, attention/persistence, sensory sensitivity; see Zentner & Bates, 2008).

Toddler Temperament Scale (TTS). The Toddler Temperament Scale (Fullard, et al., 1984; part of the CTS) was used as measure of temperament at the Time 1 evaluation. The TTS is a 97-item parent-report questionnaire that measures temperament of children ages 12-35 months. The scale is modeled after Thomas and Chess's theory of temperament (Thomas & Chess, 1977) and measures nine categories of temperament. The questionnaire asks parents to rate the frequency of their child's behavior for each item on the following scale: 1 = Almost never; 2 = Rarely; 3 = Variable, usually does not; 4 = Variable, usually does; 5 = Frequently; 6 = Almost always. Individual item scores – with reverse-scored items corrected – are then averaged to produce nine Category Scores (items per category range from eight to thirteen items). Interpretations of high and low Category Scores on the TTS (based on del Rosario et al., 2014) are summarized in Table 1. Table 1 also provides descriptions of each temperament category (as described by Thomas, Chess, & Birch, 1970) as well as sample items from each category. The TTS demonstrates adequate internal consistency (Cronbach's $\alpha \geq .7$) in our sample for the *Activity*, *Rhythmicity*, *Approach*, and *Distractibility* scales; acceptable internal consistency ($\alpha \geq .6$) for *Adaptability*, *Mood,* and *Persistence*; and poor internal consistency ($\alpha \leq .6$) for *Intensity* and *Threshold* (see Table 1).

Behavioral Style Questionnaire (BSQ). The Behavioral Style Questionnaire (McDevitt & Carey, 1996; part of the CTS) was used to measure temperament at Time 2. The BSQ is a 100-item parent-report questionnaire that measures temperament of children ages three to seven years. Like the TTS, the BSQ is based on Thomas & Chess's nine categories of temperament, and items are rated along the same six-point scale as the TTS. Category Scores on the BSQ contained a number of items that range from nine to thirteen items. Interpretations of high and low Category Scores are the same as on the TTS. Sample items from each category of the BSQ are also provided in Table 1. The BSQ demonstrates adequate internal consistency in our sample for the *Approach, Adaptability, Intensity, Mood,* and *Distractibility* scales; acceptable internal consistency for *Activity* and *Persistence;* and poor internal consistency for *Rhythmicity* and *Threshold* (see Table 1).

Mullen Scales of Early Learning (MSEL). The MSEL (Mullen, 1995) is a standardized test used to assess cognitive development in children from one-month to five years, eight months of age on five domains: gross motor, visual reception, fine motor, receptive language, and expressive language. Raw scores in each of these domains can be converted into T-scores or age equivalents. MSEL domain standard scores are internally consistent, with a median split-half reliability ranging from .75 to .85 across all ages. Inter-rater reliability is high, ranging from .91 to .99 across all age ranges (Mullen, 1995).

Childhood Autism Rating Scale (CARS). The CARS (Schopler, et al., 1980) is a 15-item clinician-rated scale measuring autism symptomology and severity that is based on clinician observation and parent report regarding the child's verbal and nonverbal communication, relating to people, imitation skills, sensory responses, as well as a general overall impression. Each of the fifteen items are scored on a scale of one to four, and then summed to produce a total score ranging from 15-60. A score of 25.5 is considered the cut-off for ASD more broadly (Chlebowski, Green, Barton, & Fein, 2010). The CARS demonstrates high internal consistency ($\alpha = .94$) and good inter-rater reliability (.71; Schopler, Reichler, & Renner, 1995).

History Form. The History Form is an investigator-developed, parent report measure with open-ended and multiple-choice questions about family demographics, pregnancy and labor, developmental, medical, and treatment history.

Modified Checklist for Autism in Toddlers (M-CHAT) / -Revised with Follow-Up (M-CHAT-R/F). The M-CHAT (Robins et al., 2001) is a parent-report, autism screening measure, consisting of 23 yes/no questions regarding a child's current behavior. Its revision (M-CHAT-R/F; Robins et al., 2009) consists of 20 yes/no questions. Higher scores indicate more ASD-related concerns. A child screens positive for ASD if he/she "fails" 3 of 23 total items, or 2 of 6 critical items on the M-CHAT, and 3 of 20 items, or 2 of 7 critical items on the M-CHAT-R. Both versions include a second-tier screening procedure involving a follow-up phone interview, on which a child screens positive if he/she fails at least two items on the follow-up. Both versions have demonstrated adequate internal reliability with the follow-up interview (Cronbach α's from .79 to .85). Positive predictive value (PPV) for ASD ranges from .68 to .79 for the M-CHAT (Kleinman et al., 2007; Robins et al., 2001; Robins, 2008). PPV is .48 for ASD and .95 for any developmental delay or concern for the M-CHAT-R/F.

Autism Diagnostic Observation Schedules (ADOS). The ADOS (Lord et al., 2000) is a semi-structured observational measure used to diagnose ASD based on behavior during activities that provide children opportunities to communicate, and engage in social interaction and pretend play. Children in the current sample were assessed with Module 1 (for children with no phrase speech) during their Time 1 evaluation, and Module 2 (for children with some phrase speech) during their Time 2 evaluation. Both modules demonstrate high inter-rater reliability (.79 and .70 for Modules 1 and 2, respectively), and acceptable to adequate internal consistency for the three ADOS domains that capture ASD symptom criteria (Lord et al., 2000).

Autism Diagnostic Interview – Revised (ADI-R). The ADI-R (Lord et al., 1994) is a semi-structured parent interview assessing ASD symptomatology based upon ICD-10 (International Statistical Classification of Diseases and Related Health Problems, Tenth Edition; World Health Organization, 1992) and DSM-IV criteria, that can be used for the diagnosis of autism for children with a mental age as low as two years (Lord et al., 1994). The ADI-R demonstrates relatively high inter-rater reliability (.63 to .89) and internal consistency (Cronbach α's from .69 to .95; Lord et al., 1994).

Toddler ASD Symptom Interview (TASI). The TASI is an investigator-developed, semi-structured parent interview of ASD symptomatology based on DSM-IV-TR criteria, and is used in place of the ADI-R for children with a mental age less than two years. The measure is currently undergoing reliability testing and validation.

Vineland Adaptive Behavior Scale, Survey Interview Form – First and Second Editions (VABS, VABS-II). The VABS (Sparrow, et al., 1984) VABS-II (Sparrow et al., 2005) are semi-structured parent interviews used to assess children's adaptive skills (what the child is currently able to do) with regards to communication, socialization, daily living skills, and motor skills. Both editions demonstrate high internal consistency and inter-rater reliability (Sparrow et al., 1984; Sparrow et al., 2005).

Data Analytic Plan

Chi-squared analyses and t-tests were used to assess whether children whose parents completed temperament measures, compared to children whose parents did not (at each time point), differed in demographic make-up and cognitive abilities. Those who returned and those who were lost to attrition by Time 2 were also compared on these factors. Chi-squared analyses and t-tests were used to check whether the ASD and non-ASD groups at both evaluation time points matched on gender composition, mean chronological age, mean verbal ratio IQ, mean nonverbal ratio IQ, race composition, and level of maternal education. (Calculation of ratio IQs are discussed in the Results section.)

Aim 1: Group differences in temperament at Time 1. Time 1 diagnostic group was coded (0 = non-ASD, 1 = ASD) and was entered as a categorical predictor into nine separate linear regressions to predict each of the nine TTS scales.

Group differences in temperament at Time 2. Time 2 diagnostic group was coded (0 = non-ASD, 1 = ASD) and entered as a categorical predictor into nine separate linear regressions to predict each of the nine BSQ scales. Due to the exploratory nature of Aim 1 analyses, and the fact that each temperament subscale represents relatively different temperament domains, criteria for statistical significance was not corrected for multiple comparisons. However, it is important to keep in mind the potential for increases in Type 1 errors as we interpret findings from these analyses.

Aim 2: Contribution of temperament beyond ASD symptomology in predicting an ASD diagnosis. To examine the extent to which temperament predicts the presence or absence of ASD at age 2, over and above a measure of autism symptomology and severity (Time 1 CARS Total Score), temperament scales (with at least acceptable internal reliability; i.e. $\alpha \geq .6$ and that were found to be significantly different by diagnostic group in Aim 1 analyses) were entered into a logistic regression with CARS Total Score to predict Time 1 diagnostic group.

Aim 3: Time 1 temperament predicting Time 2 temperament. To examine the extent to which temperament at age 2 predicts temperament at age 4, six TTS temperament category scores (those that demonstrated at least acceptable internal reliability (i.e., $\alpha \geq .6$) on both the TTS and BSQ – Activity, Adaptability, Approach, Distractibility, Mood, Persistence) were entered, simultaneously with the variable *Time Lag* (time in months between a participant's Time 1 to Time 2 evaluations), into a linear regression to predict its respective BSQ category score (e.g. TTS Activity predicting BSQ Activity). The sample used for these analyses consist of the 58 participants that had both TTS and BSQ data. Of these 58 participants, 32 maintained their ASD diagnostic group classification from Time 1 to Time 2, twenty maintained their non-ASD diagnostic group classification, four switched from ASD to non-ASD by Time 2 (i.e. lost their ASD diagnosis), and two switched from non-ASD to ASD by Time 2 (gained an ASD diagnosis). Time 1 diagnostic group was used as the designator of group membership in this sample because most of the sample retained their Time 1 diagnostic category at Time 2, which supports previous research showing diagnostic stability of diagnoses made at age 2 (Kleinman, et al., 2008).

Time 1 diagnostic group X Time 1 temperament interaction. To understand whether diagnosis at age 2 significantly moderates the extent to which Time 1 temperament predicts Time 2 temperament, each TTS category score was first centered (to help address potentially problematic high multicollinearity among predictors; Aiken & West, 1991). The centered TTS variable, Time 1 diagnostic group, and their interaction (along with Time Lag) were entered into a linear regression to predict its respective BSQ category score.

Aim 4: Gender as moderator of Time 1 and Time 2 diagnostic group differences in temperament. Gender and diagnostic group were coded (1 = Girl, 0 = Boy; 1 = non-ASD, 0 = ASD; respectively) and Aim 1 analyses were repeated with the incorporation of gender and the gender X diagnostic group interaction as predictors of temperament. Additional regressions using diagnostic group to predict temperament were run with just girls in order to understand differences in temperament between girls with and without ASD.

Power Analyses

Previous literature looking at temperamental differences between children with and without ASD report effects sizes of small to medium effects (Garon, et al., 2015), as well as medium to large effects (Clifford et al., 2013; del Rosario, et al., 2014). Post-hoc power analyses were conducted to determine how much power our data had to detect different effect sizes. Power analyses were conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). Post-hoc power analyses indicated that data from our sample of 144 participants at Time 1 had 99% power to detect medium effect sizes with linear regression, and 39% power to detect small effects. There was 99% power to detect large effects, 70% power to detect medium effects, and 14% power to detect small effects in analyses including interaction terms. Data from our sample of 73 participants at Time 2 had 90% power to detect medium effects, and 22% power to detect small effects. There was 81% power to detect large effects, and 38% power to detect medium effects, in analyses including interaction terms. Our sample of 58 participants (Aim 3) had 82% power to detect medium effects, and 18% power to detect small effects. There was 96% power to detect large effects, 65% power to detect medium effects, and 15% power to detect small effects for the interaction of diagnostic group by Time 1 temperament.

Results

Data were analyzed using an SPSS-PC package, version 23 (IBM Corp, 2013), after they were examined for entry errors, outliers, and distribution normality. Regression models were assessed for multicollinearity issues using the variance inflation factor (VIF) and tolerance statistics. A VIF of 4 or greater and a tolerance of .2 or less were used as criteria for flagging concerns about multicollinearity among predictors.

Participant characteristics

A total of 211 children failed their Time 1 screening and attended the developmental and diagnostic evaluation at Time 1 (see Figure 1). Of these 211 children, 144 had Time 1 temperament data (i.e. their parents completed the TTS). T-tests and chi-squared analyses were used to check for differences in mean age at evaluation, the breakdown of diagnostic group, gender, race, maternal education, and verbal and non-verbal ratio IQ. Histograms indicated that MSEL subscale scores were positively skewed in both groups, perhaps since the MSEL does not distinguish well among children who are lower functioning (Mullen, 1995). To help reduce the potential for violation of assumptions of parametric statistical tests, ratio IQs were calculated for each MSEL domain based on the following formula and as done previously in the literature: Ratio IQ = ([age equivalent / chronological age] * 100) (Reitzel et al., 2013; Rogers et al., 2012). Consistent with previous research (del Rosario, et al., 2014), a Verbal Ratio IQ was calculated by averaging the ratio IQs of the Receptive Language and Expressive Language domains. The Nonverbal Ratio IQ was calculated by averaging the ratio IQs of the Visual Reception and Fine Motor domains. Once transformed, scores approximated a normal distribution.

Table 2a summarizes participant characteristics of children with (n = 144) and without (n = 67) temperament data at Time 1. There were no differences in mean age at evaluation, gender, race, maternal education, verbal, and non-verbal IQ. There was a significant difference in the proportion of ASD to non-ASD diagnoses, with a higher proportion of non-ASD to ASD cases for those who did not complete the TTS, compared to those who did. Table 2b summarizes participant characteristics of children who returned for their Time 2 evaluation (n = 128), compared to those who were lost to follow-up (*n* = 83). Analyses indicated no differences in gender, race, proportion of ASD to non-ASD cases, maternal education, or Time 1 non-verbal IQ. There was a significant difference in verbal IQ at Time 1, with baby siblings who returned for their Time 2 evaluation demonstrating higher verbal IQ than those lost to follow-up.

Table 2c summarizes participant characteristics of the 73 participants with BSQ data, compared to 63 participants who did not have that data, at their Time 2 evaluation. (In addition to the 128 participants who returned for their Time 2 evaluation, 8 additional participants screened positive for the first time and attended the evaluation at Time 2). There were no significant differences in proportion of ASD to non-ASD diagnoses at Time 2, gender, or maternal education. There were significant differences in mean age at re-evaluation such that those without BSQ data were on average younger (mean = 39.7 months) than those with BSQ data (mean = 46.6 months). However, this difference is because most of the missing BSQ data came from sites that re-screened and re-evaluated baby siblings at 36 rather than 48 months (i.e. Vanderbilt and UWash during screening with the M-CHAT-R). There were also significant differences in race, with a higher ratio of White to non-White participants in the group with missing BSQ data. This finding should be interpreted with caution due to the high number of missing data on this variable.

Table 3 summarizes participant characteristics of the final samples used for Time 1 and Time 2 analyses used to address Aims 1, 2, and 4. There were no significant differences between the Time 1 ASD (*n* = 78) and non-ASD groups (*n* = 66) in gender, mean age at evaluation, race, or maternal education. The ASD group had significantly lower verbal and non-verbal ratio IQs than their peers in the non-ASD group. At Time 2, the ASD (*n* = 39) and non-ASD groups (*n* = 34) were not significantly different in mean age at evaluation or race. There was a significantly larger male to female ratio in the ASD group (4.59:1) compared to the non-ASD group (1.43:1). There were significantly more cases of no/some college and fewer advanced degrees reported for maternal education in the ASD group compared to the non-ASD group. As at Time 1, children in the ASD group had significantly lower verbal and non-verbal ratio IQs.

Table 4 summarizes participant characteristics of the sample (n = 58) used to address Aim 3. There were no differences between the ASD (n = 36) and non-ASD (n = 22) groups in gender, mean time lag in months between Time 1 and Time 2 evaluations, or maternal education. The ASD group had significantly lower Time 1 verbal and non-verbal ratio IQs, and Time 2 verbal ratio IQ. There was also a higher ratio of White to non-White participants in the ASD group compared to the non-ASD group.

Aim 1: Temperament differences between HR-ASD and HR-non-ASD baby siblings

Time 1. All TTS category scores were normally distributed. Linear regression analyses indicate that several TTS scales varied as a function of Time 1 diagnostic group (ASD versus non-ASD). Specifically, the ASD group had higher scores on Rhythmicity (B = .37, t(143) = 2.71, F(1,142) = 7.32, p = .008, R^2 = .049), indicating higher arrhythmicity; higher scores on Approach (B = .40, t(143) = *2.56*, F(1,143) = 6.56, p = .011, R^2 = .044), indicating less approach and more withdrawal; lower scores on Distractibility (B = -.85, t(142) = -5.90, F(1,143) = 34.83, p < .001, R^2 = .197), indicating less distractibility from ongoing behavior; higher scores on Adaptability (B = .34, t(143) = 2.45, F(1,142) = 6.02, p = .015, R^2 = .041), indicating decreased adaptability; and higher scores on Mood (B = .33, t(143) = 2.86, F(1,142) = 8.16, p = .005, R^2 = .054), indicating more negative mood and less positive mood. Activity (B = .17, t(143) = 1.25, F(1,142) = 1.55, p = .215, R^2 = .011), Intensity (B = .14, t(143) = 1.12, F(1,142) = 1.23, p = .264, R^2 = .009), Persistence (B = .08, t(143) = .73, F(1,142) = .54, p = .465, R^2 = .004), and

Threshold (*B* = -.10, *t*(143) = -.72, *F*(1,142) = .51, *p* = .475, R^2 = .004) did not vary as a function of Time 1 diagnostic group,. TTS scores by Time 1 diagnostic group are graphed in Figure 2.

Time 2. All BSQ category scores were normally distributed. Linear regression analyses indicate that several BSQ scales varied as a function of Time 2 diagnostic group. The ASD group had higher scores on Approach (*B* = .62, *t*(69) = 2.81, *F*(1, 68) = 7.92, *p* = .006, R^2 = .104), indicating less approach and more withdrawal; higher scores on Adaptability (*B* = .65, *t*(72) = 3.19, *F*(1,71) = 10.17, *p*= .002, R^2 = .125), indicating less adaptability; lower scores on Intensity (*B* = -.46, *t*(72) = -2.60, *F*(1,71) = 6.79, *p* = .011, R^2 = .087), indicating lower intensity reactions; lower scores on Distractibility (*B* = -.93, *t*(72) = -4.56, *F*(1,71) = 20.81, $p < .001$, R^2 = .227), indicating less distractibility; higher scores on Persistence (*B* = .43, *t*(71) = 2.73, *F*(1,70) = 7.46, *p* = .008, R^2 = .096), indicating less persistence on novel or challenging tasks; and lower scores on Threshold (*B* = -.53, *t*(72) = -3.85, *F*(1,71) = 14.86, $p < .001$, R^2 = .173), indicating lower perceptual sensitivity (requires mores stimulation). Mood (*B* = -.02, *t*(72) = -.10, *F*(1,71) = .009, *p* = .925, R^2 = 0.00), Activity (*B* = .23, *t*(72) = 1.56, *F*(1,71) = 2.45, *p* = .122, R^2 = .033), and Rhythmicity (*B* = .15, *t*(71) = .88, *F*(1,70) = .78, *p* = .382, R^2 = .011) did not vary as a function of Time 2 diagnostic group. BSQ category scores by Time 2 diagnostic group are graphed in Figure 3.

Aim 2: Predicting diagnostic group using temperament at age 2

Based on Aim 1 results, TTS category scores on Rhythmicity, Approach, Distractibility, Adaptability, and Mood were entered into a logistic regression, with Time 1 CARS Total Score, to predict diagnostic group at Time 1. Due to some missing Time 1 CARS Total Score data, 128 of 144 cases were included in this analysis. Results of the logistic regression are reported in Table 5. The overall model was significant $\chi^2(6) = 102.81$, $p < .001$, Nagelkerke R^2 = .74.

Accounting for all other predictors, lower Adaptability significantly *reduced* the likelihood of being in the ASD group. Approach and Mood approached significance as predictors, with poorer approach and increased negative mood increasing likelihood of ASD group membership, after accounting for ASD symptom severity.

Aim 3: Predicting Age 4 Temperament from Age 2 Temperament

The extent to which temperament at age 2 predicted temperament at age 4 was first examined for all baby siblings with temperament data at both time points, regardless of diagnostic group (n = 58). Results of this regression are presented in Table 6, and depicted visually in Figure 4. Average time lag in months between the first and second evaluations was 23.6 months (SD = 5.9), with a range of 11 to 39 months. After controlling for time lag, Time 1 scores predicted a significant amount of variance in Time 2 scores on the temperament categories of Approach, Adaptability, Mood, Persistence, and Distractibility. Time 1 Activity did not predict a significant amount of variance in Time 2 Activity.

Given the range in time lag between evaluations, post-hoc moderation analyses were conducted to evaluate whether time lag moderated the extent to which Time 1 temperament predicted Time 2 temperament. Time lag and Time 1 temperament variables were centered, their interaction calculated, and entered into a linear regression to predict Time 2 temperament. Table 7 summarizes the results of these moderation analyses. The stability of Adaptability, Mood, and Distractibility from Time 1 to Time 2 was significantly moderated by time lag. Results from simple slopes analyses used to follow-up on significant interactions indicated that Time 1 Adaptability, Mood, and Distractibility predicted a significant amount of variance in their respective Time 2 scores, but only for low (-1 SD below the mean) and average levels of time lag. In other words, as the lag between evaluation time points extended beyond the mean 23.6

months, Time 1 temperament was less predictive of Time 2 temperament along these domains. Conditional effects of time lag on stability of temperament are summarized in Table 8, and depicted visually in Figure 5.

Time 1 diagnostic group X Time 1 temperament interaction. A linear regression (Time Lag, Centered TTS score, Time 1 Diagnostic Group, Centered TTS X Time 1 Diagnostic Group interaction predicting BSQ score) was used to assess whether the relationship between Time 1 and Time 2 temperament was moderated by presence/absence of ASD. Time 1 diagnosis did not significantly moderate the relationship between Time 1 and Time 2 scores for all six temperament categories assessed, after accounting for time lag. Table 9 and Figure 6 summarize these results.

Correlations between Time 1 and Time 2 Temperament. Given our sample does not have sufficient power to detect small to medium effects in the above moderation analysis, we wanted to further explore potential differences depicted in Figure 6 between diagnostic groups in the strength of association of each TTS category score with its respective BSQ category. Pearson's correlation coefficients were calculated between each TTS and its respective BSQ score for Time 1 ASD and non-ASD groups separately. Table 10 presents these results. Correlations between Time 1 and Time 2 scores are significant for the following scales, but only for the ASD group: Approach, Adaptability, Mood, Distractibility, and Persistence. Compared to the moderate correlations for the ASD group, Time 1 and Time 2 temperament were weakly associated and correlations were non-significant for the non-ASD group. Activity was not significantly correlated for either group.

Aim 4: Gender moderating relationship between diagnostic group and temperament

Due to low power to detect even medium-sized effects when investigating interactions within our Time 2 sample, gender difference analyses were conducted with our Time 1 sample only. Within the Time 1 sample (n = 144), there were 95 boys (50 [52.6%] with ASD, 45 [47.4%] without ASD) and 49 girls (28 [57.1%] with ASD, 21 [42.9%] without ASD). Results from linear regressions indicate several findings regarding the effects of gender, Time 1 diagnostic group, and their interaction on Time 1 temperament (see Tables 11 and 12, and Figure 7). Gender only moderated the relationship between diagnostic group and Persistence. Boys with ASD were rated as significantly less persistent on challenging tasks than boys without ASD, a pattern not found in girls. Boys with ASD were also rated as significantly less rhythmic, less distractible, and with greater negative mood than non-ASD boys. Girls with ASD were rated as significantly less approaching than non-ASD girls and ASD boys.

Discussion

This study investigated several questions regarding temperament in a sample of HR baby siblings who screened positive on an autism-specific screener, and who therefore demonstrated higher risk for ASD (not only genetic risk but also presence of parent-reported concerns about ASD) than most samples of HR baby siblings investigated thus far.

Time 1 (Age 2) Temperament Differences

First, we were interested in comparing parent-rated temperament of baby siblings with ASD to those without ASD. As expected, at age 2 HR-ASD baby siblings were less rhythmic (less regular) in eating, elimination, and sleep-wake patterns; less approaching towards novel objects, situations, and people; less responsive to external stimuli and therefore less distractible from ongoing behavior; less adaptable to changes in the environment (including physical experiences such as being dressed or washed, as well as adapting to and learning new rules); and displayed more frequent negative rather than positive mood, compared to HR-non-ASD baby siblings. However, we found no significant differences in activity level, intensity of emotional expressions, persistence on challenging tasks, and perceptual sensitivity (threshold). It may be the case that increased activity in children with ASD is more apparent later in childhood or young adulthood, as observed in previous literature (del Rosario et al., 2014; Brock et al., 2012). Persistence also did not distinguish between groups, perhaps because many children in our non-ASD group had a developmental delay that may make persisting on new and challenging tasks more difficult (Baker et al., 2003; Redding & Morgan, 1988). Children with ASD may be more perceptually sensitive to certain stimuli (e.g. non-social stimuli and objects), but not to others (social stimuli), which may have contributed to poor internal consistency of the scale as well as the null overall difference in perceptual sensitivity compared to the non-ASD group.

In all, these results at Time 1 largely replicate previous findings about temperament profiles observed during toddlerhood in HR baby siblings who go on to develop ASD, and provide further evidence that temperament may help distinguish young children with and without ASD. The most robust difference between HR-ASD and HR-non-ASD siblings was on distractibility, likely because items on this scale tap into difficulties in attunement to one’s environment and increased inflexibility of attentional focus, which are often part of ASD symptomology, and likely most apparent to parents.

Time 2 (Age 4) Temperament Differences

We were also interested in temperament differences between HR-ASD and HR-non-ASD siblings at age 4, which has not been previously explored in the literature. We hypothesized that, despite effects of maturation, temperamental differences observed in toddlerhood between children with and without ASD would largely be maintained as they enter preschool, potentially influencing how adults perceive behavior of and interact with HR-ASD siblings. Indeed, at age 4 HR-ASD siblings were found to be less approaching toward novel events, objects, and people; less adaptable; less distractible from ongoing behavior; and less persistent compared to HR-non-ASD siblings. Rather than increased intensity of emotions, we found that our HR-ASD group were slightly less intense in emotional expression at age 4. The mixed findings in this literature may be due to fact that atypical affective expressivity manifests in a range of ways in children with ASD, from blunted, extreme, inappropriate, to poorly differentiated (Bieberich & Morgan, 2004; Lickel, MacLean, Blakeley-Smith, & Hepburn, 2012; Loveland, 2005; Mazefsky et al., 2013).

We also found that these children had lower perceptual sensitivity (require more stimulation to evoke a response), although it is important to note the lower internal consistency of the BSQ Threshold scale. Unlike at age 2, baby siblings with and without ASD did not differ in mood, activity level, or rhythmicity at age 4. Rhythmicity had poorer internal reliability on the BSQ. The lack of difference in mood may be attributed to the lower intensity of emotional expressions observed in our HR-ASD sample, such that the increased negative mood observed in toddlerhood was less apparent to parents at age 4. It may also be the case that HR-ASD siblings are expressing distress in other ways not readily observable to parents. While our study did not have sufficient data on participation in early intervention services to allow us to account for treatment effects in these analyses, it is possible that children with ASD showing increased negative mood, activity level, and arrhythmicity at age 2 no longer demonstrated those increases as a result of treatment.

Contribution of Temperament in Predicting Presence or Absence of ASD

After accounting for Time 1 CARS Total Score and other temperament features, children with higher negative mood were still more likely to be diagnosed with ASD (over three times more likely for every increase in one unit on the TTS Mood scale) at age 2. In support of our hypothesis, this finding suggests that information related to emotional expression and regulation, while not core symptoms necessary for a diagnosis of ASD, may provide important additional information about risk for ASD. These findings are in line with recent literature suggesting that more nuanced analyses of CARS scores (e.g. looking at an Emotional Reactivity factor) may facilitate understanding of diverse symptom profiles in toddlers with ASD (Moulton, Bradbury, Barton, & Fein, 2016). Our results also indicate that children who were less approaching were more likely to be diagnosed with ASD at age 2, after accounting for their CARS score and other temperament features (almost two times more likely per increase in one unit on the Approach scale). Despite a certain degree of overlap between the ASD symptom cluster of atypical social interaction and communication and the temperament construct of Approach (e.g. children who struggle to initiate social contact will also be less approaching towards novel objects or people), information about temperament provided additional benefit in predicting risk for ASD. It may be the case that at the mean CARS total score in this sample (i.e. 28.17, which is above the score of 25.5 broadly found to be the cutoff for autism; Chlebowski, et al., 2010), children who were rated to be least approaching and most withdrawn were more likely to be in the ASD group compared to peers who may demonstrate subclinical ASD symptoms.

Interestingly, children who were less adaptable at age 2 were *less* likely to be diagnosed with ASD, after accounting for CARS score and other temperament features. It may be the case that at the mean CARS total score of 28.17 (which indicates mild autism severity), and average levels of rhythmicity, approach, mood, and distractibility (i.e. the child does not display any extreme temperament patterns), poor adaptability to change may better reflect typical behavior in toddlerhood or the effects of other developmental concerns (i.e. cognitive or language delay) rather than ASD. Additionally, this finding may suggest that for children who generally demonstrate this “mild” presentation (i.e. mild ASD symptoms and balanced temperament profiles), behavioral rigidity is less predictive of ASD.

Predicting Age 4 Temperament from Age 2 Temperament

As expected, we found that as a whole group, temperament of HR baby siblings at age 2 predicted a significant amount of variance in temperament at age 4. HR baby siblings who were less adaptive, less approaching, less distractible, had more negative mood, and were less persistent when they were two-years-old were displaying similar behavioral tendencies at age 4, suggesting that – as with findings with TD samples – some temperament domains are relatively stable from toddlerhood to preschool. Relative to scores on other domains, scores on Distractibility was most consistent from Time 1 to Time 2, suggesting that a child’s tendency to attune to environmental stimuli and to shift attention away from ongoing activities may be least subject to change over time, or least targeted in early intervention. Activity was not consistent over time. For some children in this group, activity levels decreased by age 4, which may reflect normative development of behavioral regulation skills due to maturation and schooling. However, this pattern was not observed for all HR baby siblings.

It is important to note that the stability of Distractibility (as well as that of Adaptability and Mood) was shown to decrease over time. This may reflect children displaying different behavioral patterns as they start preschool, where they engage with a structured learning environment and increased opportunities to interact with new adults and peers. Alternatively, it is possible that compared to TD samples, temperament is a less stable construct for HR baby siblings (or less well measured using tools developed for TD children). Finally, parental ratings of temperament may be less consistent as time passes.

We did not find that Time 1 diagnostic group moderated the relationship between Time 1 and Time 2 temperament. However, correlational analyses by diagnostic group indicated that several temperament domains were significantly associated only within the ASD group. Some research suggest that stability of temperament in TD children is often moderated by multiple factors such as parenting behaviors and cognitive ability (Rubin, Burgess, & Hastings, 2002; Stifter et al., 2008). The low correlation between Time 1 and Time 2 temperament in the non-ASD group may reflect the contributions of such moderating variables, and this moderation effect may not be the same for children with ASD (e.g. parenting behaviors may not influence the stability of temperament for children with ASD as it would for TD children). It is likely that differences by diagnostic group in the relationship between Time 1 and Time 2 temperament are small effects, not readily detected by our moderation analysis given our sample size. However, it may be fruitful for future studies using larger samples to explore whether and why stability of temperament differs depending on the presence or absence of autism. For example, parents of children with ASD may view their child's behavior as less susceptible to change over time, and may be more likely to continue rating their child on the extremes of temperament scales. Alternatively, autism symptoms may negatively impact the development of self-regulation skills important for improvement in adaptability to change, inhibition of negative affect during distress, and persistence on challenging activities.

Gender Differences in Temperament at Age 2

While we did not detect significant moderation by gender on the relationship between diagnosis and temperament, our findings do point to some gender-specific differences in temperament between toddlers with and without ASD. Several potential interpretations may explain these findings. First, certain domains of temperament may manifest differently in boys and girls with ASD (or interact with the emergence of ASD symptoms in different ways), such that girls look more similar to their non-ASD female peers in terms of their rhythmicity, persistence, distractibility, and mood, but more dissimilar in approach behaviors, relative to the comparison of boys with and without ASD. Second, temperament (in addition to indicating risk for ASD) may also indicate protective factors against risk for ASD in HR-non-ASD baby siblings (Garon et al, 2009). If being more rhythmic, more persistent, more distractible, and having more positive mood distinguishes siblings without ASD from those with ASD, our findings may suggest that these protective factors are present for boys, but not girls. Third, these findings may reflect the influence of gender norms on parental perceptions of “normative” versus “atypical” behavioral patterns for boys versus girls. That is, girls with ASD may be – at comparable levels as boys with ASD – less rhythmic, less persistent, less distractible, and have greater negative mood than non-ASD peers, but these differences may be less readily observable to parents, or may be downplayed as “typical” behavior for a girl. Furthermore, parents may expect daughters to be more approaching towards (rather than withdrawn from) novel people, objects, and situations, compared to sons. Thus, low approach tendencies may be rated as more extreme for girls. Fourth, our smaller sample of girls may not have had enough power to detect some differences observed in our sample of boys.

Clinical Implications

Our study builds upon previous research on temperament in HR baby sibling cohorts, providing further evidence of differences in early behavioral patterns that distinguish HR siblings who develop ASD from those who do not. Importantly, we replicate several previous findings regarding temperamental differences between HR-ASD and HR-non-ASD siblings, but within a sibling cohort that shared not only high genetic risk for autism, but who all displayed behaviors that led to a positive screen on an autism-specific screener. In this regard, our study shows that temperament may serve as an additional behavioral marker of risk for ASD, beyond ASD symptom-related concerns. We further demonstrate that certain domains of temperament (particularly Approach and Mood) may provide additional utility for predicting risk for autism at age 2, beyond ASD symptom severity.

Our study also sheds light on how temperament manifests in HR baby siblings with and without ASD at the start of preschool, a critical transition period in which children begin to spend more time away from primary caregivers, have increased opportunities to interact with same-aged peers, and engage with novel activities, people, and situations. A child with ASD who is viewed as less adaptable and less persistent on challenging tasks, for example, may evoke unique responses from adults working with him (beyond the influence of being identified as the child in the classroom with ASD) that impacts his experience at school (e.g. teachers may feel less encouraged to challenge this child to learn to adapt to and engage with novel activities.)

Our findings regarding the stability of temperament from age 2 to 4 further speaks to the above point. While age 2 temperament predicted significant variance in age 4 temperament for the whole HR baby sibling sample, we found that parents were more consistent in their ratings for their child with ASD, perhaps reflecting perceptions of the pervasiveness and persistence of behavioral patterns associated with an autism diagnosis. Again, this may have unintended and unrecognized influences on the ways in which parents and other adults may interact with a child with ASD. Understanding which behavioral patterns may be least subject to change over time (e.g. poor distractibility) may also shed light on areas that require increased intervention for children with ASD.

Lastly, our findings about gender differences suggest that there may be important ways in which early patterns of behavior indicative of emerging ASD manifest differently for boys and girls. Specifically, previous findings about increased arrhythmicity, negative mood, poor distractibility, and poor persistence in HR-ASD baby siblings may apply mostly to boys. Early temperamental patterns associated with emerging ASD may look differently (or may be subtler) for girls, and at present, we may not have an adequate understanding of what those patterns are. Alternatively, parental perceptions of normative behavior for boys and girls may influence parent-reported temperament in ways that "down-play" (or increase concern about) the atypicality of certain behavioral patterns. We know from previous research that social interaction concerns predictive of an ASD diagnosis are more often voiced by parents of boys than by parents of girls (Little, Wallisch, Salley, & Jamison, 2016). Furthermore, girls with ASD may engage in compensatory social behaviors that tend to "camouflage" social deficits (e.g. staying in close proximity to peers) that makes it more difficult for adults to pick up on atypical social interaction patterns (Dean, Harwood, & Kasari, 2016). As our field continues to investigate the role of temperament as an early behavioral marker of ASD, and researchers begin to think more deeply about the ways in which our knowledge about autism may not have adequately captured how ASD manifests in girls, we need to continue exploring (rather than controlling for) the effects of gender in our research.

Limitations and Future Directions

Several challenges identified in previous literature regarding research on temperament in HR baby siblings are also limitations in our study. First, our earliest measurement of temperament was at age 2, which limits our ability to conclude whether temperament profiles associated with ASD preceded or emerged in parallel with autism symptoms. Even in studies that measured temperament in HR baby siblings as early as six months, it is difficult to provide a definitive answer to this question (Bryson et al., 2007). Second, our study continues to rely on parent-report of temperament. While this allows us to make interesting hypotheses about how parental perceptions may influence our conclusions about temperament in ASD (e.g. perceptions about gender norms), it is also difficult to identify and parse out potential biases that parents bring to the table, and the extent to which their ratings reflect parental perceptions versus biologically determined behavior patterns. This is particularly interesting among parents of HR baby siblings, who have previous experience with a child diagnosed with ASD. Parental perceptions of temperament in their younger child may be influenced by increased awareness of atypical behavior patterns (parents of HR baby siblings tend to voice concerns about their child's development earlier than parents without an older child with ASD; Herlihy, Knoch, Vibert, & Fein, 2015; Ozonoff et al., 2009). Furthermore, parents may be rating their younger child's temperament while using their older child's temperament as a baseline, and it is unclear whether or how this baseline would be different for parents without an older child with ASD. Parents of HR baby siblings may be interacting with this younger child in unique ways that influence the development of temperamental styles. Growing up with an older sibling with ASD may also influence temperament in unique ways that we do not yet understand (Clifford et al., 2013). Indeed, this question of how generalizable results from studies with HR baby sibling cohorts are of the general population is one that continues to be heavily debated in this literature (Clifford et al., 2013). Direct measures of temperament (e.g. clinician ratings based on behavioral observations, paradigms – such as a frustration paradigm – designed to evoke behavioral responses) can provide additional information about temperament without being subject to parental biases.

The significantly higher proportion of non-ASD to ASD cases among the group who did *not* complete the TTS at Time 1 may be interpreted in a few ways. It may be the case that parents whose children received a non-ASD diagnosis entered the evaluation with fewer concerns about their child, and were thus less invested in completing all measures. This issue does raise the question of whether temperament data from our non-ASD comparison group comes from a self-selective group of slightly more motivated parents, or parents who had more concerns about their child's temperament (more likely to provide more extreme ratings nearer to those reported by the ASD group) compared to parents to opted out of completing the TTS.

Our study compared temperament between two groups that were highly similar – not only matched on most demographic factors, but also in terms of genetic risk for autism as well as presence of identified concerns about ASD by parents (i.e. positive screen for autism). While this allows us to better isolate temperament for investigation, there are several questions that our study could not adequately address due to our lack of a low-risk baby sibling group. For example, we cannot speak to how temperament profiles associated with ASD may be tied to genetic risk for ASD. Previous research (Zwaigenbaum et al., 2005) found no differences between HR-non-ASD baby siblings and low-risk siblings in temperament within the first two years of life, and it would have been interesting to see if we could have replicated these findings in our sample and at the preschool age.

In addition to questions about whether temperament profiles precede, parallel, or follow the emergence of ASD, there is still much to learn about whether and how temperament profiles observed in our findings are unique to risk for ASD. Researchers have been particularly interested in understanding converging pathways of risk for ASD and ADHD, and how early temperament provides information about risk for either or both disorders (Grefer, Flory, Cornish, Hatton, & Roberts, 2016; Miller, Iosif, Young, Hill, & Ozonoff, 2016; Sizoo, van der Gaag, & van den Brink, 2015). Furthermore, if temperament plays a key factor in the heterogeneity of the presentation of ASD, it will be important for future studies to explore how temperament varies within the ASD group. Future research should also investigate how temperament predicts risk for comorbid psychopathology in children with ASD later in life.

While the TTS and BSQ used to measure temperament in this study allowed us to measure a wide range of temperament domains, these measures (and Thomas and Chess's theory of temperament on which these measures were based) have been criticized because they conceptualize temperament in a manner that arbitrarily divorces temperament from motivation or capacity for behavior. It would be important to think about how motivational deficits in ASD (e.g. social motivation; Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012) may influence temperament, and how our measures may not adequately capture this caveat. This raises a further question of whether temperament in ASD is best conceptualized using models and best measured using tools developed for TD populations (De Pauw, 2011; Shiner et al, 2003). Other researchers have also raised the issue that positive and negative mood are not mutually exclusive, as is assumed on the Mood scales of the TTS and BSQ. Furthermore, it would be fruitful to have more nuanced understanding of what consists of "increased negative mood" (e.g. is it increased irritability, sadness, and/or tantruming?).

Cross-cultural research on temperament has suggested that parents from different cultural backgrounds may conceptualize temperament domains in different ways, which may influence parental ratings of temperament (Gartstein, Knyazev, & Slobodskaya, 2005; Gartstein, et al., 2006). For example, in a study of temperament in TD 6-7-year-old children, Chinese parents associated smiling and laughing behavior in their child more closely with the temperament domain of *surgency* (conceptualized as high positive affect, approach behaviors, as well as high activity and impulsivity), while American parents conceptualized the same behavior as closely related to effortful control (attentional, emotional, and behavioral control; Ahadi, Rothbart, & Ye, 1993). The intersectionality of cultural differences and perceptions about gender norms may also interact in unique ways that influence parental ratings of temperament. The lack of racial and ethnic diversity in our sample highlights the important need for researchers to take more active steps to recruit and retain ethnically and racially diverse samples in order to provide opportunities to investigate cultural considerations important for the conceptualization of temperament in ASD.

Conclusions

While temperament is only one of several early behavioral risk markers for ASD (Sacrey et al., 2015), it continues to be an important source of information that facilitates early detection of ASD and understanding of course and heterogeneity of presentations of ASD. This study demonstrated that certain domains of temperament can distinguish HR baby siblings with and without ASD, and potentially in different ways for boys and girls. Furthermore, measures of temperament may account for areas of behavior not well captured by standard autism diagnostic tools. Continuing to investigate how temperament interacts with development of foundational learning skills, emerging ASD symptoms, and the child’s environment to predict risk for ASD will be important for future research.

References

Adamek, L., Nichols, S., Tetenbaum, S. P., Bregman, J., Ponzio, C. A., & Carr, E. G. (2011). Individual temperament and problem behavior in children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 26*(3), 173-183.

Ahadi, S. A., Rothbart, M. K., & Ye, R. (1993). Children's temperament in the US and China: Similarities and differences. *European Journal of Personality, 7*(5), 359-378.

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

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

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC.

Bailey, A., Palferman, S., Heavey, L., & Le Couteur, A. (1998). Autism: The phenotype in relatives. *Journal of Autism and Developmental Disorders*, *28*, 369–392.

Bailey Jr, D. B., Hatton, D. D., Mesibov, G., Ament, N., & Skinner, M. (2000). Early development, temperament, and functional impairment in autism and fragile X syndrome. *Journal of Autism and Developmental Disorders*, *30*(1), 49-59.

Baio, J. (2014). Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators; Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveill. Summ*, *63*, 1-21.

Baker, B. L., McIntyre, L. L., Blacher, J., Crnic, K., Edelbrock, C., & Low, C. (2003). Pre‐school children with and without developmental delay: behaviour problems and parenting stress over time. *Journal of Intellectual Disability Research, 47*(4‐5), 217-230.

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

Bieberich, A. A., & Morgan, S. B. (2004). Self-regulation and affective expression during play in children with autism or Down syndrome: A short-term longitudinal study. Journal of *Autism and Developmental Disorders, 34*(4), 439-448.

Brock, M. E., Freuler, A., Baranek, G. T., Watson, L. R., Poe, M. D., & Sabatino, A. (2012). Temperament and sensory features of children with autism. *Journal of Autism and Developmental Disorders*, *42*(11), 2271-2284.

Bryson, S. E., Zwaigenbaum, L., Brian, J., Roberts, W., Szatmari, P., Rombough, V., & McDermott, C. (2007). A prospective case series of high-risk infants who developed autism. *Journal of Autism and Developmental Disorders*, *37*(1), 12-24.

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.

Chlebowski, C., Green, J. A., Barton, M. L., & Fein, D. (2010). Using the childhood autism rating scale to diagnose autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*(7), 787-799.

Cicchetti, D. E., & Cohen, D. J. (1995). *Developmental psychopathology, Vol. 1: Theory and methods.* John Wiley & Sons.

Clifford, S. M., Hudry, K., Elsabbagh, M., Charman, T., Johnson, M. H., & BASIS Team. (2013). Temperament in the first 2 years of life in infants at high-risk for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(3), 673-686.

Daniels, A. M., & Mandell, D. S. (2014). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism*, *18*(5) 583 –597.

Dean, M., Harwood, R., & Kasari, C. (2016). The art of camouflage: Gender differences in the social behaviors of girls and boys with autism spectrum disorder. *Autism.*

del Rosario, M., Gillespie-Lynch, K., Johnson, S., Sigman, M., & Hutman, T. (2014). Parent-reported temperament trajectories among infant siblings of children with autism. *Journal of Autism and Developmental Disorders*, *44*(2), 381-393.

De Pauw, S. S., & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry & Human Development, 41*(3), 313-329.

De Pauw, S. S., Mervielde, I., Van Leeuwen, K. G., & De Clercq, B. J. (2011). How temperament and personality contribute to the maladjustment of children with autism. *Journal of Autism and Developmental Disorders*, *41*(2), 196-212.

Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(8), 788-797.

Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: a meta-analysis. *Psychological Bulletin, 132*(1), 33.

Falck-Ytter, T., Bölte, S., & Gredebäck, G. (2013). Eye tracking in early autism research. *Journal of Neurodevelopmental Disorders, 5*(1), 1.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Fullard, W., McDevitt, S. C., & Carey, W. B. (1984). Assessing temperament in one-to three-year-old children. *Journal of Pediatric Psychology*, *9*(2), 205-217.

Garon, N., Bryson, S. E., Zwaigenbaum, L., Smith, I. M., Brian, J., Roberts, W., & Szatmari, P. (2009). Temperament and its relationship to autistic symptoms in a high-risk infant sib cohort. *Journal of Abnormal Child Psychology*, *37*(1), 59-78.

Garon, N., Zwaigenbaum, L., Bryson, S., Smith, I. M., Brian, J., Roncadin, C., ... & Roberts, W. (2015). Temperament and its Association with Autism Symptoms in a High-risk Population. *Journal of Abnormal Child Psychology*, *44*(4), 757-769.

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant Behavior and Development, 26*(1), 64-86.

Gartstein, M. A., Gonzalez, C., Carranza, J. A., Ahadi, S. A., Ye, R., Rothbart, M. K., & Yang, S. W. (2006). Studying cross-cultural differences in the development of infant temperament: People's Republic of China, the United States of America, and Spain. *Child Psychiatry and Human Development*, *37*(2), 145-161.

Gartstein, M. A., Knyazev, G. G., & Slobodskaya, H. R. (2005). Cross-cultural differences in the structure of infant temperament: United States of America (US) and Russia. *Infant Behavior and Development*, *28*(1), 54-61.

Grefer, M., Flory, K., Cornish, K., Hatton, D., & Roberts, J. (2016). The emergence and stability of attention deficit hyperactivity disorder in boys with fragile X syndrome. *Journal of Intellectual Disability Research, 60*(2), 167-178.

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., ... & McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. *Child Development, 58*(2), 505-529.

Gould, J., & Ashton-Smith, J. (2011). Missed diagnosis or misdiagnosis? Girls and women on the autism spectrum. *Good Autism Practice (GAP)*, *12*(1), 34-41.

Hepburn, S. L., & Stone, W. L. (2006). Using Carey Temperament Scales to assess behavioral style in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(5), 637-642.

Herlihy, L., Knoch, K., Vibert, B., & Fein, D. (2015). Parents' first concerns about toddlers with autism spectrum disorder: Effect of sibling status. *Autism*, *19*(1), 20-28.

IBM Corp (2013). *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.

Kasari, C., Sigman, M., Mundy, P., & Yirmiya, N. (1988). Caregiver interactions with autistic children. *Journal of Abnormal Child Psychology*, *16*(1), 45-56.

Kiff, C. J., Lengua, L. J., & Zalewski, M. (2011). Nature and nurturing: Parenting in the context of child temperament. *Clinical Child and Family Psychology Review, 14*(3), 251-301.

Kleinman, J. M., Robins, D. L., Ventola, P. E., Pandey, J., Boorstein, H. C., Esser, E. L., … Fein, D. (2007). The modified checklist for autism in toddlers: a follow-up study investigating the early detection of autism spectrum disorders. *J Autism Dev Disord*, *38*(5), 827–839.

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S., ... & Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(4), 606-615.

Kochanska, G., Murray, K.T., & Harlan, E.T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*(2), 220-232.

Konstantareas, M. M., & Stewart, K. (2006). Affect regulation and temperament in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *36*(2), 143-154.

Landa, R., & Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: a prospective study. *Journal of Child Psychology and Psychiatry*, *47*(6), 629-638.

Lickel, A., MacLean, W. E., Blakeley-Smith, A., & Hepburn, S. (2012). Assessment of the prerequisite skills for cognitive behavioral therapy in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*(6), 992-1000.

Little, L. M., Wallisch, A., Salley, B., & Jamison, R. (2016). Do early caregiver concerns differ for girls with autism spectrum disorders? *Autism*, 1362361316664188.

Mazefsky, C. A., Herrington, J., Siegel, M., Scarpa, A., Maddox, B. B., Scahill, L., & White, S. W. (2013). The role of emotion regulation in autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(7), 679-688.

Moore, V., & Goodson, S. (2003). How well does early diagnosis of autism stand the test of time? Follow-up study of children assessed for autism at age 2 and development of an early diagnostic service. *Autism*, *7*(1), 47-63.

Moulton, E., Bradbury, K., Barton, M., & Fein, D. (2016). Factor Analysis of the Childhood Autism Rating Scale in a Sample of Two Year Olds with an Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders,* 1-14.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord*, *24*(5), 659–685.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*, *30*(3), 205–223.

McDevitt, S. C., & Carey, W. B. (1996). *Manual for the Behavioral Style Questionnaire*. Scottsdale, AZ: Behavioral-Developmental Initiatives.

Miller, M., Iosif, A. M., Young, G. S., Hill, M. M., & Ozonoff, S. (2016). Early Detection of ADHD: Insights From Infant Siblings of Children With Autism. *Journal of Clinical Child & Adolescent Psychology*, 1-8.

Mullen, E. M. (1995). *Mullen Scales of Early Learning: Manual*. Minneapolis: NCS Pearson, Inc.

Mundy, P. C., Henderson, H. A., Inge, A. P., & Coman, D. C. (2007). The modifier model of autism and social development in higher functioning children. *Research and practice for Persons with Severe Disabilities: The Journal of TASH*, *32*(2), 124.

Ostfeld-Etzion, S., Feldman, R., Hirschler-Guttenberg, Y., Laor, N., & Golan, O. (2015). Self-regulated compliance in preschoolers with autism spectrum disorder: The role of temperament and parental disciplinary style. *Autism*, 1362361315615467.

Ozonoff, S., Young, G. S., Steinfeld, M. B., Hill, M. M., Cook, I., Hutman, T., ... & Sigman, M. (2009). How early do parent concerns predict later autism diagnosis? *Journal of Developmental and Behavioral Pediatrics*, *30*(5), 367.

Paterson, G., & Sanson, A. (1999). The association of behavioural adjustment to temperament, parenting and family characteristics among 5-year-old children. *Social Development, 8*(3), 293-309.

Pedlow, R., Sanson, A., Prior, M., & Oberklaid, F. (1993). Stability of maternally reported temperament from infancy to 8 years. *Developmental Psychology*, *29*(6), 998-1007.

Redding, R. E., Morgan, G. A., & Harmon, R. J. (1988). Mastery motivation in infants and toddlers: Is it greatest when tasks are moderately challenging?. *Infant Behavior and Development*, *11*(4), 419-430.

Reitzel, J., Summers, J., Lorv, B., Szatmari, P., Zwaigenbaum, L., Georgiades, S., & Duku, E. (2013). Pilot randomized controlled trial of a Functional Behavior Skills Training program for young children with Autism Spectrum Disorder who have significant early learning skill impairments and their families. *Research in Autism Spectrum Disorders, 7*(11), 1418–1432.

Robins, D., Fein, D., Barton, M. & Green, J. (2001) The Modified Checklist for Autism in Toddlers: An initial study investing the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *31*(2), 131-144.

Robins, D. L. (2008). Screening for autism spectrum disorders in primary care settings. *Autism*, *12*(5), 537–556.

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

Rogers, S. J., Estes, A., Lord, C., Vismara, L., Winter, J., Fitzpatrick, A., … Dawson, G. (2012). Effects of a brief Early Start Denver model (ESDM)-based parent intervention on toddlers at risk for autism spectrum disorders: a randomized controlled trial. *J Am Acad Child Psy*, *51*(10), 1052–1065.

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development*, 569-578.

Rothbart, M. K., & Bates, J. E. (1998). *Temperament: Handbook of Child Psychology*. John Wiley & Sons.

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: origins and outcomes. *Journal of Personality and Social Psychology*, *78*(1), 122-135.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, *72*(5), 1394-1408.

Rubin, K. H., Burgess, K. B., & Hastings, P. D. (2002). Stability and social–behavioral consequences of toddlers' inhibited temperament and parenting behaviors. *Child development*, *73*(2), 483-495.

Sacrey, L. A. R., Bennett, J. A., & Zwaigenbaum, L. (2015). Early infant development and intervention for autism spectrum disorder. *Journal of Child Neurology*, *30*(14), 1921-1929.

Salley, B., Miller, A., & Bell, M. A. (2012). Associations between temperament and social responsiveness in young children. *Infant and Child Development*, *22*(3), 270-288.

Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development, 13*(1), 142-170.

Sanson, A., Pedlow, R., Cann, W., Prior, M., & Oberklaid, F. (1996). Shyness ratings: Stability and correlates in early childhood. *International Journal of Behavioral Development*, 19, 705-724.

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders*, *10*(1), 91-103.

Schopler, E., Reichler, R. J., & Renner, B. R. (1995). *Childhood Autism Rating Scale: Manual*. Los Angeles, CA: Western Psychological Services.

Shiner, R., & Caspi, A. (2003). Personality differences in childhood and adolescence: Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry*, *44*(1), 2-32.

Sizoo, B. B., van der Gaag, R. J., & van den Brink, W. (2015). Temperament and character as endophenotype in adults with autism spectrum disorders or attention deficit/hyperactivity disorder. *Autism*, *19*(4), 400-408.

Sparrow, S. S., Balla, D. A., Cicchetti, D. V., Harrison, P. L., & Doll, E. A. (1984). *Vineland Adaptive Behavior Scales.* Circle Pines, MN: American Guidance Service.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales, Second Edition*. Circle Pines, MN: AGS Publishing.

Stifter, C. A., Putnam, S., & Jahromi, L. (2008). Exuberant and inhibited toddlers: Stability of temperament and risk for problem behavior. *Development and Psychopathology*, *20*(02), 401-421.

Szatmari, P., Chawarska, K., Dawson, G., Georgiades, S., Landa, R., Lord, C., ...& Halladay, A. (2016). Prospective Longitudinal Studies of Infant Siblings of Children with Autism: Lessons Learned and Future Directions. *Journal of the American Academy of Child & Adolescent Psychiatry*, *5*5(3), 179-187.

Thomas, A., & Chess, S. (1977). Temperament and development. New York: Brunner/Mazel Publishers.

Todd, J. T., & Dixon, W. E. (2010). Temperament moderates responsiveness to joint attention in 11-month-old infants. *Infant Behavior and Development, 33*(3), 297-308.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

Zentner, M., & Bates, J. E. (2008). Child temperament: An integrative review of concepts, research programs, and measures. *International Journal of Developmental Science*, *2*(1-2), 7-37.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, *23*(2), 143-152.

Table 1

Toddler Temperament Scale (TTS) and Behavioral Style Questionnaire (BSQ) temperament categories, internal reliability, and category score interpretation

Temperament Category	Example items	Cronbach's α	Lower Category Score (closer to 1)	Higher Category Score (closer to 6)
Activity: level of motor activity	TTS: The child fidgets during quiet activities (e.g. story-time). BSQ: The child speaks so quickly that it's difficult to understand him/her.	TTS: $\alpha = .75^{\dagger}$ BSQ: $\alpha = .67^{\text{₸}}$	Less active	More active
Rhythmicity: degree of regularity of eating, elimination and the sleep-wake cycle	TTS: The child gets sleepy at about the same time each evening. BSQ: The child eats about the same amount at supper from day to day.	TTS: $\alpha = .76^{\dagger}$ BSQ: $\alpha = .56$	Rhythmic	Arrhythmic
Approach: tendency to approach or withdraw in response to novel objects or people	TTS: The child's initial reaction to a new babysitter is rejection (crying, clinging to mother, etc.) BSQ: The child had trouble leaving the mother the first three days when he/she entered school.	TTS: $\alpha = .86^{\dagger}$ BSQ: $\alpha = .81^{\dagger}$	Approaching	Withdrawn/ cautious
Adaptability: pace in adapting to changes in the environment	TTS: The child is still wary of strangers after 15 minutes. BSQ: The child is slow to adjust to changes in household rules.	TTS: $\alpha = .67^{\text{₸}}$ BSQ: $\alpha = .84^{\dagger}$	Quick	Gradual
Intensity: intensity of emotional response (positive or negative)	TTS: The child vigorously resists additional food or milk when full (spits out, claps mouth closed, bats at spoon, etc.)	TTS: $\alpha = .58$ BSQ: $\alpha = .70^{\dagger}$	Mild	Intense

	BSQ: The child reacts strongly (cries or complains) to a disappointment or failure.			
Mood: general mood or "disposition"	TTS: The child remains pleasant when hungry and waiting for food to be prepared. BSQ: The child becomes angry with one of his/her playmates.	TTS: α = .69[₸] BSQ: α = .74[†]	More positive, less negative	More negative, less positive
Persistence: ability to stay with an activity despite external distraction	TTS: The child pays attention to games with parent for only a minute or so. BSQ: The child is reluctant to give up when trying to do a difficult task.	TTS: α = .62[₸] BSQ: α = .67[₸]	Persistent	Non-persistent
Distractibility: degree to which behavior is altered by external stimuli	TTS: The child ignores voices when playing with a favorite toy. BSQ: The child interrupts an activity to listen to conversation around him/her.	TTS: α = .86[†] BSQ: α = .84[†]	Behavior rarely altered by external stimuli	Behavior often altered by external stimuli
Threshold: intensity of stimulation required to evoke a discernable response	TTS: The child ignores differences in taste or consistency of familiar foods. BSQ: The child becomes upset or cries over minor falls or bumps.	TTS: α = .59 BSQ: α = .32	Low perceptual sensitivity (more stimulation needed)	High perceptual sensitivity (less stimulation needed)

Note: [†] Subscale demonstrates adequate internal reliability; α ≥ .7.
[₸] Subscale demonstrates acceptable internal reliability; α = .6 to .7.

Table 2a

Characteristics of participants with and without Time 1 temperament data (TTS) at Time 1 evaluation

	With TTS	Without TTS	Test of group difference
N	144	67	
Gender *N* (*% male*)	95 (66.0%)	40 (59.7%)	$\chi^2(1) = .78$, $p = .377$
Chronological age in months *M (SD)*	23.0 (4.3)	23.5 (3.7)	$t(206) = .81$, $p = .421$
Diagnostic Group *N (%)*			
ASD	78 (54.2%)	23 (34.3%)	$\chi^2(1) = 7.21$, $p = .007$
Non-ASD	66 (45.8%)	44 (65.7%)	
Verbal ratio IQ *M (SD)*[a]	69.46 (27.50)	77.20 (32.00)	$t(197) = 1.73$, $p = .086$
Nonverbal ratio IQ *M (SD)*[b]	87.90 (19.44)	89.97 (26.45)	$t(198) = .54$, $p = .590$
Race *N (% White)*[c]	114 (82.6%)	18 (75.0%)	$\chi^2(1) = .78$, $p = .376$
Maternal Education *N (%)*[d]			$\chi^2(2) = 1.04$, $p = .594$
No/Some College	41 (41.4%)	24 (39.3%)	
Bachelor's Degree	35 (35.4%)	26 (42.6%)	
Advanced Degree	23 (23.2%)	11 (18.0%)	

[a] Results on Verbal Ratio IQ are based on data from 140 participants with TTS data, and 59 participants without TTS data. Ratio IQs have a mean of 100, and SD of 15.

[b] Results on Nonverbal Ratio IQ are based on data from 141 participants with TTS data, and 59 participants without TTS data. Ratio IQs have a mean of 100, and SD of 15.

[c] Results on Race are based on data from 138 participants with TTS data, and 24 participants without TTS data.

[d] Results on Maternal Education are based on data from 99 participants with TTS data, and 61 participants without TTS data. Maternal education of No diploma, High School Diploma, GED, Associate's Degree, Some College, and Vocational/Technical Degree were collapsed into the category "No/Some College". Master's, Doctorate, and other Advanced Degrees were collapsed into the category "Advanced Degree".

Table 2b

Characteristics of participants who returned versus lost to Time 2 evaluation

	Attended Time 2	Did not attend Time 2	Test of group difference
N	128	83	
Gender *N* (*%* male)	82 (64.1%)	53 (63.9%)	$\chi^2(1) = .001$, $p = .976$
Diagnostic Group *N (%)*			
ASD	59 (46.1%)	40 (48.2%)	$\chi^2(1) = .09$, p = .765
Non-ASD	69 (53.9%)	43 (51.8%)	
Time 1 Verbal ratio IQ *M (SD)*[a]	75.63 (28.75)	64.68 (27.63)	$t(192) = -2.59$, $p = .010$, $d = .39$
Time 1 Nonverbal ratio IQ *M (SD)*[b]	89.88 (22.68)	85.56 (19.69)	$t(194) = -1.35$, $p = .180$
Race *N* (*%* White)[c]	63 (85.1%)	36 (75.0%)	$\chi^2(1) = 1.96$, $p = .162$
Maternal Education *N* (*%*)[d]			$\chi^2(2) = 2.72$, $p = .257$
No/Some College	37 (38.9%)	25 (53.2%)	
Bachelor's Degree	37 (38.9%)	15 (31.9%)	
Advanced Degree	21 (22.1%)	7 (14.9%)	

[a] Results on Verbal Ratio IQ are based on data from 123 participants who returned for Time 2, and 71 participants lost to Time 2. Ratio IQs have a mean of 100, and SD of 15.

[b] Results on Nonverbal Ratio IQ are based on data from 124 participants who returned for Time 2, and 72 participants lost to Time 2. Ratio IQs have a mean of 100, and SD of 15.

[c] Results on Race are based on data from 74 participants who returned for Time 2, and 48 participants lost to Time 2.

[d] Results on Maternal Education are based on data from 95 participants who returned for Time 2, and 47 participants lost to Time 2.

Table 2c

Characteristics of participants with and without Time 2 temperament data (BSQ) at Time 2 evaluation

	With BSQ	Without BSQ	Test of group difference
N	73	63	
Gender *N* (*%* male)	52 (71.2%)	37 (58.7%)	$\chi^2(1) = 2.34$, $p = .126$
Chronological age in months *M (SD)*	46.6 (7.2)	39.7 (9.3)	$t(134) = -4.89$, $p < .001$, $d = .83$
Diagnostic Group *N* (*%*)			
ASD	39 (53.4%)	35 (55.6%)	$\chi^2(1) = .06$, $p = .804$
Non-ASD	34 (46.6%)	28 (44.4%)	
Verbal ratio IQ *M (SD)*[a]	81.60 (26.38)	75.95 (24.10)	$t(72) = -.79$, $p = .432$
Nonverbal ratio IQ *M (SD)*[a]	85.98 (24.04)	90.84 (24.30)	$t(72) = .73$, $p = .467$
Race *N* (*%* White)[b]	65 (91.5%)	22 (66.7%)	$\chi^2(1) = 10.20$, $p = .001$
Maternal Education *N* (*%*)[c]			$\chi^2(2) = .65$, $p = .652$
No/Some College	17 (29.8%)	23 (45.1%)	
Bachelor's Degree	23 (40.4%)	21 (41.2%)	
Advanced Degree	17 (29.8%)	7 (13.7%)	

[a] Results on Verbal Ratio IQ and Nonverbal Ratio IQ are based on data from 57 participants who had BSQ data, and 17 participants who did not BSQ data at Time 2. Ratio IQs have a mean of 100, and SD of 15.

[b] Results on Race are based on data from 71 participants who had BSQ data, and 33 participants who did not BSQ data at Time 2.

[c] Results on Maternal Education are based on data from 57 participants who had BSQ data, and 51 participants who did not BSQ data at Time 2.

Table 3

Sample characteristics of Time 1 and Time 2 analyses

	ASD	non-ASD	Test of group difference
Time 1			
N	78	66	
Gender *N* (% male)	50 (64.1%)	45 (68.2%)	$\chi^2(1) = .27, p = .607$
Chronological age in months *M (SD)*	23.3 (4.2)	22.3 (4.5)	$t(142) = 1.35, p = .179$
Verbal ratio IQ *M (SD)*[a]	57.26 (18.83)	81.19 (25.52)	$t(138) = -6.36, p < .001$, $d = 1.07$
Nonverbal ratio IQ *M (SD)*[b]	81.72 (15.46)	93.06 (15.95)	$t(139) = -4.28, p < .001$, $d = .72$
Race *N* (% White)[c]	61 (83.6%)	53 (81.5%)	$\chi^2(1) = .10, p = .754$
Maternal Education (*N*, %)[d]			$\chi^2(2) = 1.41, p = .493$
No/Some College	22 (46.8%)	19 (36.5%)	
Bachelor's Degree	14 (29.8%)	21 (40.4%)	
Advanced Degree	11 (23.4%)	12 (23.1%)	
Time 2			
N	39	34	
Gender *N* (% male)	32 (82.1%)	20 (58.8%)	$\chi^2(1) = 4.78, p = .029$
Chronological age in months *M (SD)*	45.5 (6.8)	48.8 (7.7)	$t(71) = -1.96, p = .054$
Verbal ratio IQ *M (SD)*[e]	72.68 (29.16)	93.95 (15.44)	$t(55) = -3.56, p = .001$, $d = .91$
Nonverbal ratio IQ *M (SD)*[e]	79.38 (25.43)	95.12 (18.91)	$t(55) = -2.68, p = .010$, $d = .70$
Race *N* (% White)[f]	36 (94.7%)	29 (87.9%)	$\chi^2(1) = 1.07, p = .300$

Maternal Education *N (%)*[g]			$\chi^2(2) = 6.02$, $p = .049$
No/Some College	12 (40.0%)	5 (18.5%)	
Bachelor's Degree	13 (42.3%)	10 (37.0%)	
Advanced Degree	5 (16.7%)	12 (44.4%)	

[a] Results on Time 1 Verbal Ratio IQ are based on data from 74 participants in the ASD group, and 66 participants in the non-ASD group.

[b] Results on Time 1 Nonverbal Ratio IQ are based on data from 75 participants in the ASD group, and 66 participants in the non-ASD group.

[c] Results on Time 1 Race are based on data from 73 participants in the ASD group, and 65 participants in the non-ASD group.

[d] Results on Time 1 Maternal Education are based on data from 47 participants in the ASD group, and 52 participants in the non-ASD group.

[e] Results on Time 2 Verbal Ratio IQ and Nonverbal Ratio IQ are based on data from 33 participants in the ASD group, and 24 participants in the non-ASD group.

[f] Results on Time 2 Race are based on data from 38 participants in the ASD group, and 33 participants in the non-ASD group.

[g] Results on Time 2 Maternal Education are based on data from 30 participants in the ASD group, and 27 participants in the non-ASD group.

Table 4

Sample characteristics of participants with both Time 1 and Time 2 temperament data (n = 58)

	ASD (at Time 1)	non-ASD (at Time 1)	Test of group difference
N	36	22	
Gender *N* (% male)	28 (77.8%)	16 (72.7%)	$\chi^2(1) = .19$, $p = .663$
Time Lag (months, *SD*)	22.70 (5.9)	25.1 (5.6)	$t(56) = -1.53$, $p = .132$
Time 1 Verbal ratio IQ *M (SD)*[a]	61.96 (20.10)	85.26 (24.80)	$t(54) = -3.86$, $p < .001$, $d = 1.03$
Time 1 Nonverbal ratio IQ *M (SD)*[a]	81.21 (14.12)	91.45 (14.14)	$t(55) = -2.66$, $p = .010$. $d = .72$
Time 2 Verbal ratio IQ *M (SD)*[b]	73.05 (28.73)	93.47 (21.51)	$t(48) = -2.76$, $p = .010$, $d = .80$
Time 2 Nonverbal ratio IQ *M (SD)*[b]	78.76 (24.51)	91.43 (22.32)	$t(48) = -1.83$, $p = .073$
Race *N* (% White)[c]	34 (97.1%)	17 (77.3%)	$\chi^2(1) = 5.66$, $p = .017$
Maternal Education *N* (%)[d]			$\chi^2(2) = 3.69$, $p = .158$
No/Some College	10 (38.5%)	2 (11.8%)	
Bachelor's Degree	9 (34.6%)	9 (52.9%)	
Advanced Degree	7 (26.9%)	6 (35.3%)	

[a] Results on Time 1 Verbal Ratio IQ are based on data from 34 participants in the ASD group. Results on Time 1 Nonverbal Ratio IQ are based on data from 35 participants in the ASD group. Ratio IQs have a mean of 100, and SD of 15.

[b] Results on Time 2 Verbal Ratio IQ and Time 2 Nonverbal Ratio IQ are based on data from 31 participants in the ASD group, and 19 participants in the non-ASD group.

[c] Results on Race are based on data from 35 participants in the ASD group.

[d] Results on Maternal Education are based on data from 26 participants in the ASD group, and 17 participants in the non-ASD group.

Table 5

Logistic Regression: Predicting Time 1 Diagnostic Group (n = 128)

Predictor	*B*	*Exp(B)* [95% CI]	*p*
Time 1 CARS Total Score	.48	1.61 [1.35, 1.93]	< .001
Temperament			
Rhythmicity	.17	1.18 [.50, 2.76]	.696
Approach	.68	1.97 [.98, 3.94]	.056
Adaptability	-1.40	.25 [.06, .97]	.046
Mood	1.14	3.16 [.92, 10.66]	.067
Distractibility	-.53	.59 [.26, 1.36]	.213

Note: Diagnostic Group coded 1=ASD, 0=non-ASD

Table 6

Multiple Linear Regression: Time 1 Temperament Predicting Time 2 Temperament (n = 58)

Variable	β	t	p	$F(2, 55)$	p	R^2
Activity				1.58	.215	.05
Time Lag	-.22	-1.66	.103			
TTS Activity	.09	.69	.494			
Approach				4.53	.015	.14
Time Lag	-.17	-1.31	.195			
TTS Approach	.34	2.69	.010			
Adaptability				4.86	.011	.15
Time Lag	-.21	-1.67	.100			
TTS Adaptability	.33	2.68	.010			
Mood				6.30	.003	.19
Time Lag	-.27	-2.21	.032			
TTS Mood	.36	2.95	.005			
Persistence				4.21	.020	.13
Time Lag	<.01	.004	.997			
TTS Persistence	.36	2.87	.006			
Distractibility				7.79	.001	.23
Time Lag	-.04	-.33	.742			
TTS Distractibility	.47	3.90	<.001			

Note: Outcome variable for each regression is the TTS category score's respective BSQ category score. TTS category scores here are *not* centered.

Table 7

Multiple Linear Regression: Time 1 Temperament Predicting Time 2 Temperament, moderated by Time Lag (n = 58)

Variable	β	t	p	$F(3, 54)$	p	R^2
Activity				2.25	.093	.11
Time Lag	-.25	-1.91	.061			
TTS Activity	.09	.73	.471			
TTS Activity X Time Lag	-.24	-1.86	.068			
Approach				3.92	.013	.18
Time Lag	-.11	-.83	.409			
TTS Approach	.32	2.60	.012			
TTS Approach X Time Lag	-.20	-1.57	.124			
Adaptability				7.58	<.001	.30
Time Lag	-.13	-1.09	.281			
TTS Adaptability	.33	2.91	.005			
TTS Adaptability X Time Lag	-.39	-3.35	.001			
Mood				7.88	<.001	.30
Time Lag	-.14	-1.14	.261			
TTS Mood	.31	2.64	.011			
TTS Mood X Time Lag	-.37	-3.03	.004			
Persistence				3.20	.031	.15
Time Lag	.03	.25	.802			
TTS Persistence	.36	2.83	.007			
TTS Persistence X Time Lag	-.37	-1.07	.287			
Distractibility				7.34	<.001	.29
Time Lag	-.11	-.88	.384			
TTS Distractibility	.42	3.56	.001			
TTS Distractibility X Time Lag	-.27	-2.21	.031			

Note: Outcome variable for each regression is the TTS category score's respective BSQ category score. TTS category scores and Time Lag are centered.

Table 8

Conditional Effects of Time Lag on Stability of Temperament

Variable	*B*	*SE*	*β*	*t*	*p*
Adaptability					
-1 SD Time Lag	.82	.19	.67	4.40	<.001
Mean Time Lag	.40	.14	.33	2.91	.005
+1 SD Time Lag	-.01	.19	-.01	-.05	.965
Mood					
-1 SD Time Lag	.75	.17	.63	4.35	<.001
Mean Time Lag	.36	.14	-.14	2.64	.011
+1 SD Time Lag	-.03	.20	-.02	-.14	.893
Distractibility					
-1 SD Time Lag	.72	.16	.65	4.57	<.001
Mean Time Lag	.47	.13	.42	3.56	.001
+1 SD Time Lag	.21	.19	.19	1.12	.270

Table 9

Multiple Linear Regression Results: Time 1 Temperament Predicting Time 2 Temperament, moderated by diagnostic group (n = 58)

Variable	β	t	p	$F(4,53)$	p	R^2
Activity				1.36	.261	.09
Time Lag	-.19	-1.42	.162			
TTS Activity	-.11	-.51	.614			
Dx	.06	.42	.679			
TTS Activity X Dx	.28	1.328	.190			
Approach				3.38	.016	.21
Time Lag	-.14	-1.09	.280			
TTS Approach	.04	.16	.878			
Dx	.20	1.55	.126			
TTS Approach X Dx	.33	1.39	.169			
Adaptability				4.58	.003	.26
Time Lag	-.14	-1.17	.249			
TTS Adaptability	.14	.75	.457			
Dx	.25	2.00	.051			
TTS Adaptability X Dx	.29	1.60	.115			
Mood				3.717	.010	.22
Time Lag	-.26	-2.08	.042			
TTS Mood	.16	.77	.447			
Dx	-.13	-1.02	.312			
TTS Mood X Dx	.25	1.21	.232			
Persistence				3.77	.009	.22
Time Lag	.08	.61	.542			
TTS Persistence	.05	.21	.835			
Dx	.19	1.49	.142			
TTS Persistence X Dx	.47	1.66	.103			
Distractibility				5.43	.001	.29
Time Lag	-.16	-1.22	.229			
TTS Distractibility	.14	.66	.511			
Dx	.21	-1.71	.094			
TTS Distractibility X Dx	.33	1.59	.118			

Note: Outcome variable for each regression is the TTS category score's respective BSQ category score. TTS category scores here are centered variables. "Dx" = Time 1 Diagnostic Group (coded 1=ASD)

Table 10

Correlation Between Time 1 and Time 2 Temperament, By Time 1 Diagnostic Group (n = 58)

		Time 1 Temperament					
Time 1 Diagnostic Group	Time 2 Temperament	Approach	Adaptability	Mood	Distractibility	Persistence	Activity
ASD	Approach	.451**					
	Adaptability		.542**				
	Mood			.476**			
	Distractibility				.476**		
	Persistence					.541**	
	Activity						.250

		Time 1 Temperament					
Time 1 Diagnostic Group	Time 2 Temperament	Approach	Adaptability	Mood	Distractibility	Persistence	Activity
Non-ASD	Approach	.040					
	Adaptability		.131				
	Mood			.077			
	Distractibility				.274		
	Persistence					.068	
	Activity						-.124

** Correlation is significant at the 0.01 level (2-tailed).

Table 11

Linear Regression Results: Moderation of Relationship Between Time 1 Diagnostic Group and Temperament by Gender (n = 144)

Variable	*B*	*SE*	*β*	*t*	*p*	*F*(3, 140)	*p*	R^2
Activity						1.02	.387	.02
Dx	-.24	.17	-.14	-1.4	.160			
Gender	-.23	.19	-.14	-1.20	.232			
Dx X Gender	.18	.29	.08	.61	.541			
Rhythmicity						3.37	.020	.07
Dx	-.48	.17	-.28	-2.83	.005			
Gender	.03	.19	.02	.14	.889			
Dx X Gender	.33	.29	.14	1.14	.256			
Approach						3.73	.013	.07
Dx	-.24	.19	-.13	-1.26	.209			
Gender	.46	.22	.23	2.12	.036			
Dx X Gender	-.44	.33	-.17	-1.35	.180			
Adaptability						2.06	.109	.04
Dx	-.30	.17	-.18	-1.74	.085			
Gender	.09	.20	.05	.44	.657			
Dx X Gender	-.12	.29	-.05	-.42	.678			
Intensity						.864	.461	.02
Dx	-.04	.16	-.03	-.26	.795			
Gender	.18	.18	.11	1.00	.320			
Dx X Gender	-.29	.27	-.14	-1.10	.272			
Mood						4.84	.003	.09
Dx	-.37	.14	-.26	-2.64	.009			
Gender	.22	.16	.15	1.36	.175			
Dx X Gender	.15	.24	.08	.63	.530			
Persistence						2.50	.062	.05
Dx	-.28	.14	-.21	-2.03	.045			
Gender	-.17	.16	-.12	-1.05	.298			
Dx X Gender	.60	.24	.31	2.50	.014			
Distractibility						12.55	<.001	.21
Dx	1.01	.18	.53	5.73	<.001			
Gender	.17	.20	.08	.82	.414			
Dx X Gender	-.49	.31	-.18	-1.59	.113			

Threshold						.51	.673	.01
Dx	.20	.18	.12	1.15	.253			
Gender	.18	.20	.10	.89	.373			
Dx X Gender	-.28	.30	-.12	-.96	.340			

Note: "Dx" = Diagnosis. Gender = Effect of being a girl, when Dx = ASD. Dx: Effect of having *not* having ASD, when Gender = Boy. Gender X Dx = Effect of gender on the relationship between Dx and temperament.

Table 12

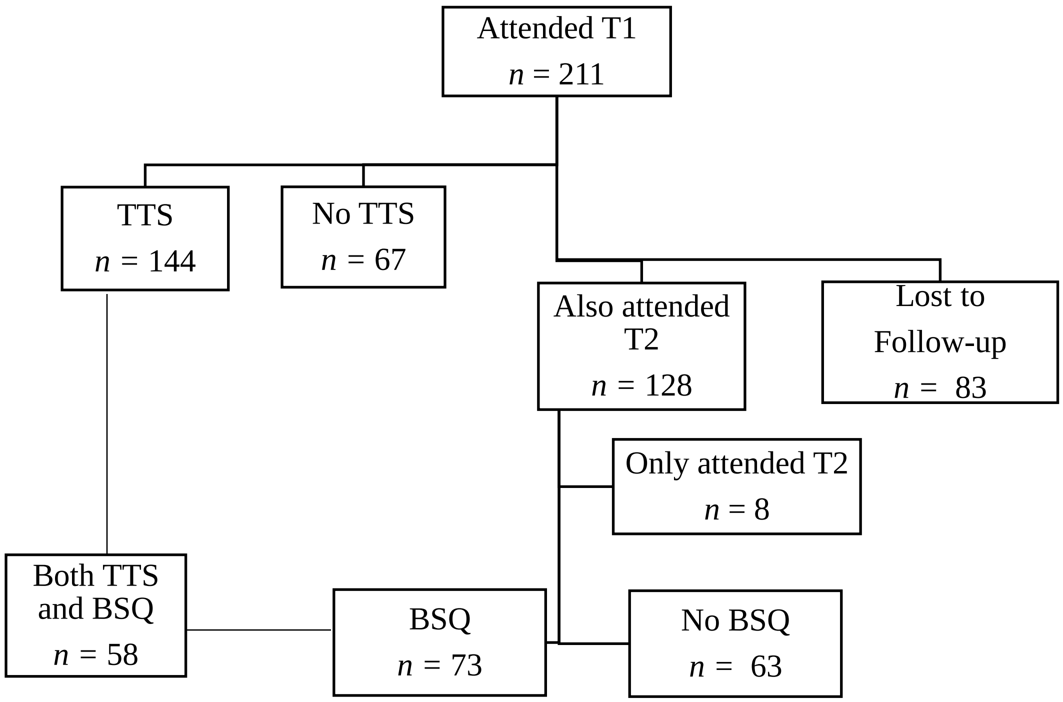
Linear Regression Results: Time 1 Diagnostic Group Predicting Temperament in Girls (n = 49)

Variable	*B*	*SE*	*β*	*t*	*p*	*F*(1, 47)	R^2
Activity	.06	.24	-.04	.25	.803	.06	<.01
Rhythmicity	.15	.25	.09	.59	.557	.35	.09
Approach	.68	.28	.34	2.46	.018	6.04	.11
Adaptability	.42	.25	.24	1.66	.104	2.75	.06
Intensity	.33	.21	.23	1.59	.118	2.53	.05
Mood	.22	.21	.15	1.05	.301	1.09	.02
Persistence	-.32	.17	-.26	-1.87	.067	3.50	.07
Distractibility	-.53	.27	-.28	-1.98	.054	3.90	.08
Threshold	-.09	.24	.05	.35	.727	.12	<.01

Note: Diagnostic Group coded 1= ASD, 0=non-ASD

Figure 1

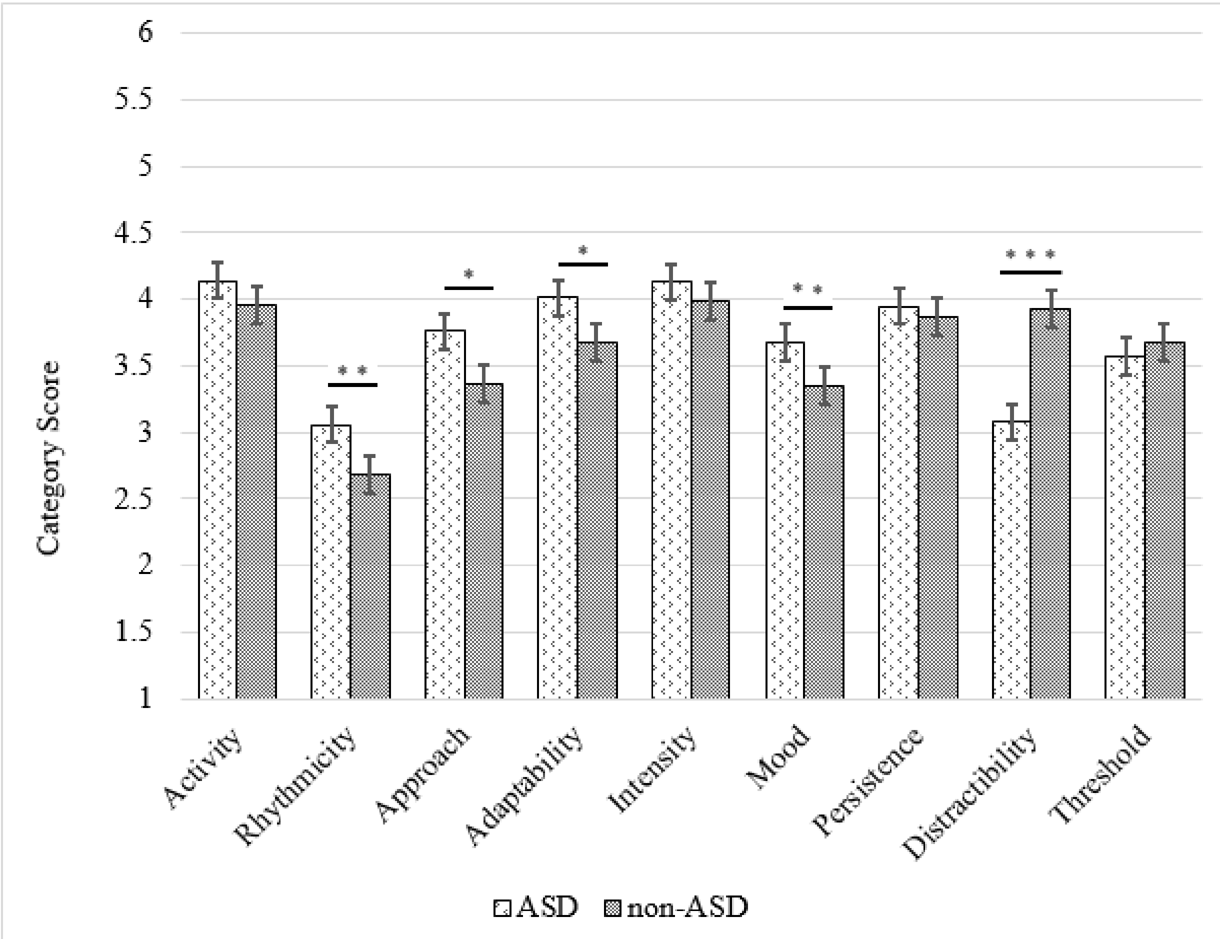
Flow of Participants



Note: T1 = Time 1 Evaluation; T2 = Time 2 Evaluation; TTS = Toddler Temperament Scale; BSQ = Behavioral Style Questionnaire

Figure 2

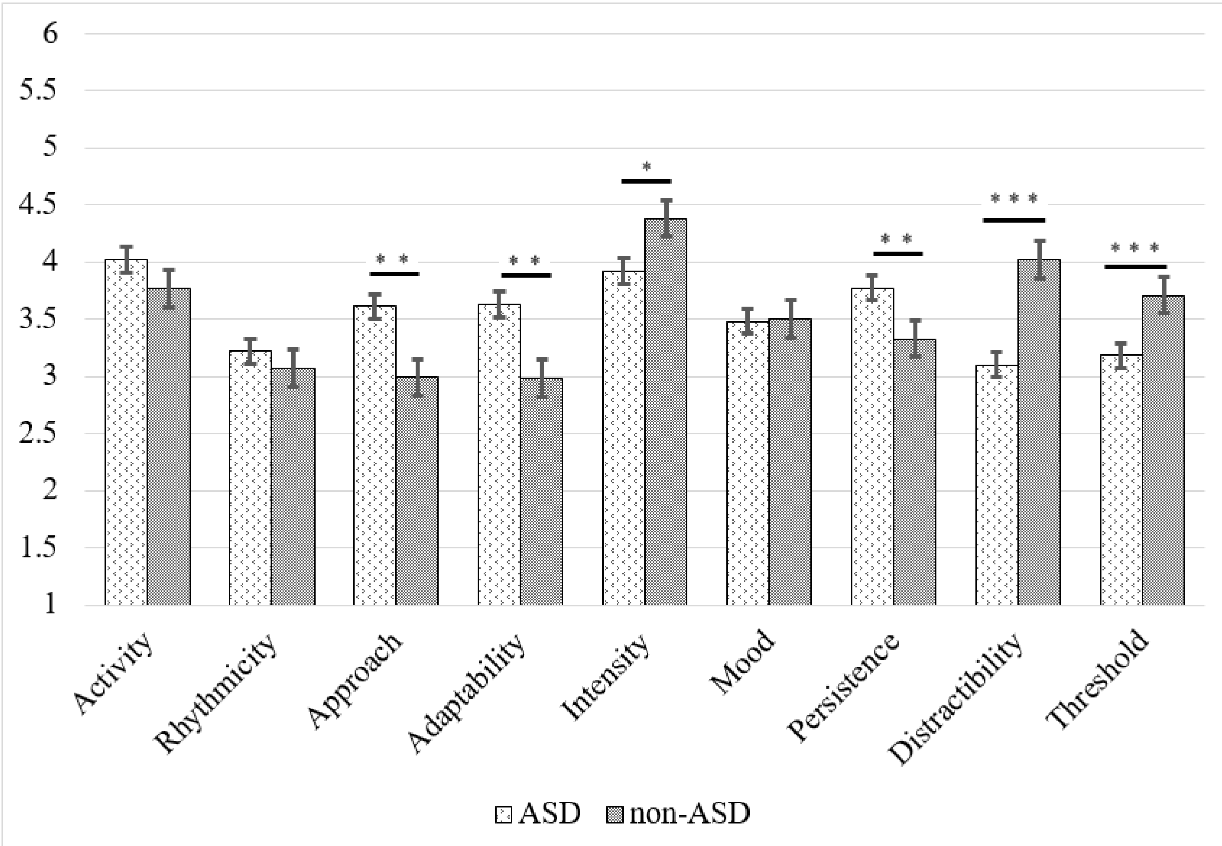
TTS Category Scores by Time 1 Diagnostic Group



Note: Error Bars: +/-2 SE; * p < .05, ** p < .01, *** p < .001

Figure 3

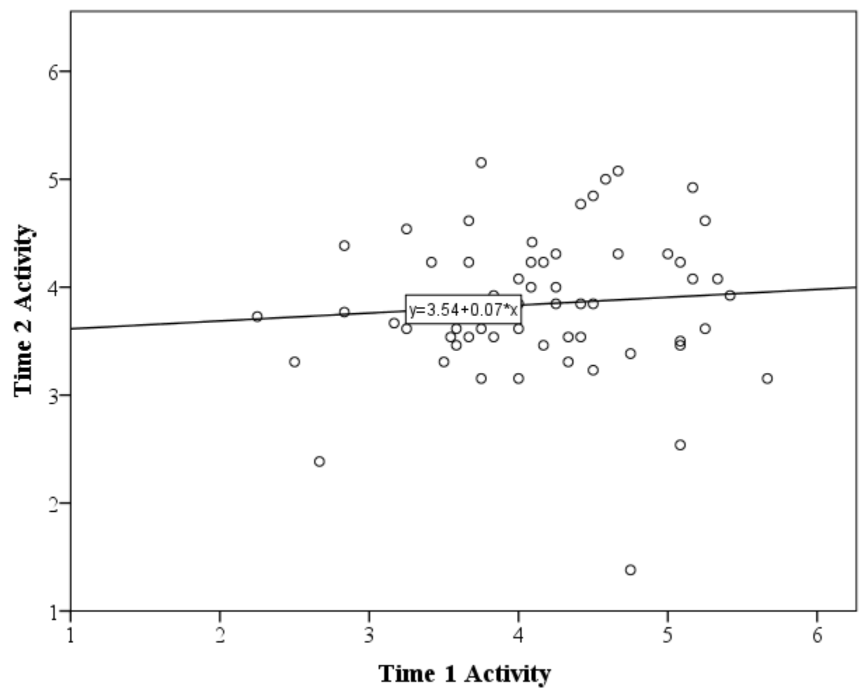
BSQ Category Scores by Time 2 Diagnostic Group

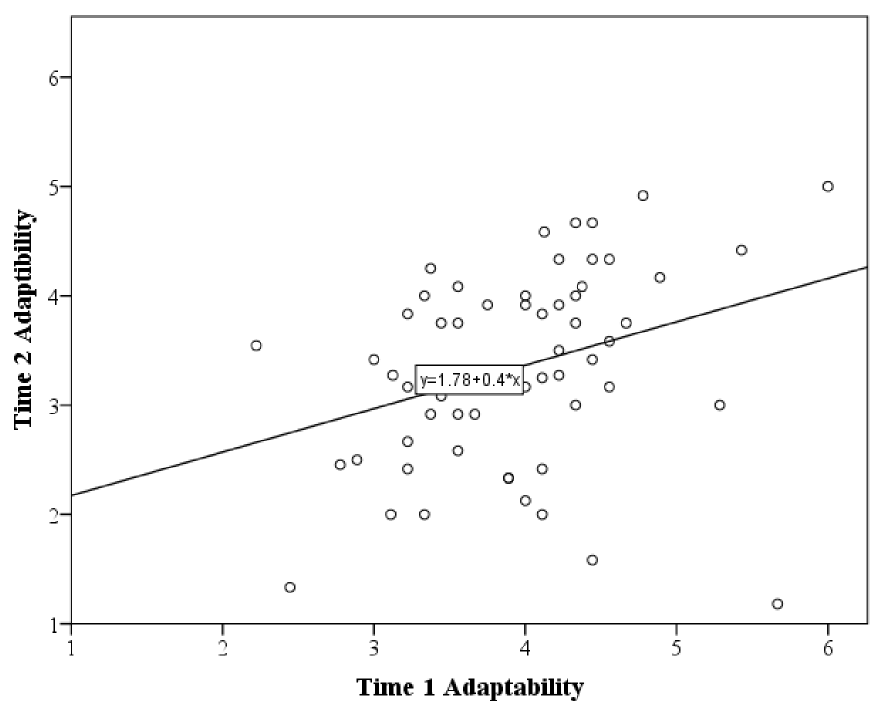


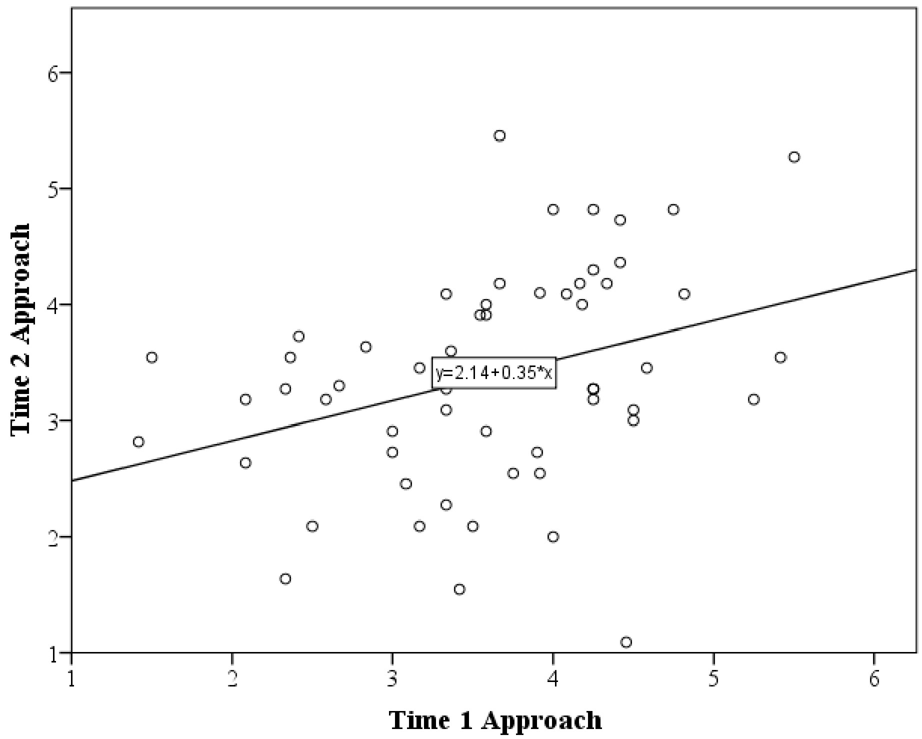
Note: Error Bars: +/-2 SE; * $p < .05$, ** $p < .01$, *** $p < .001$

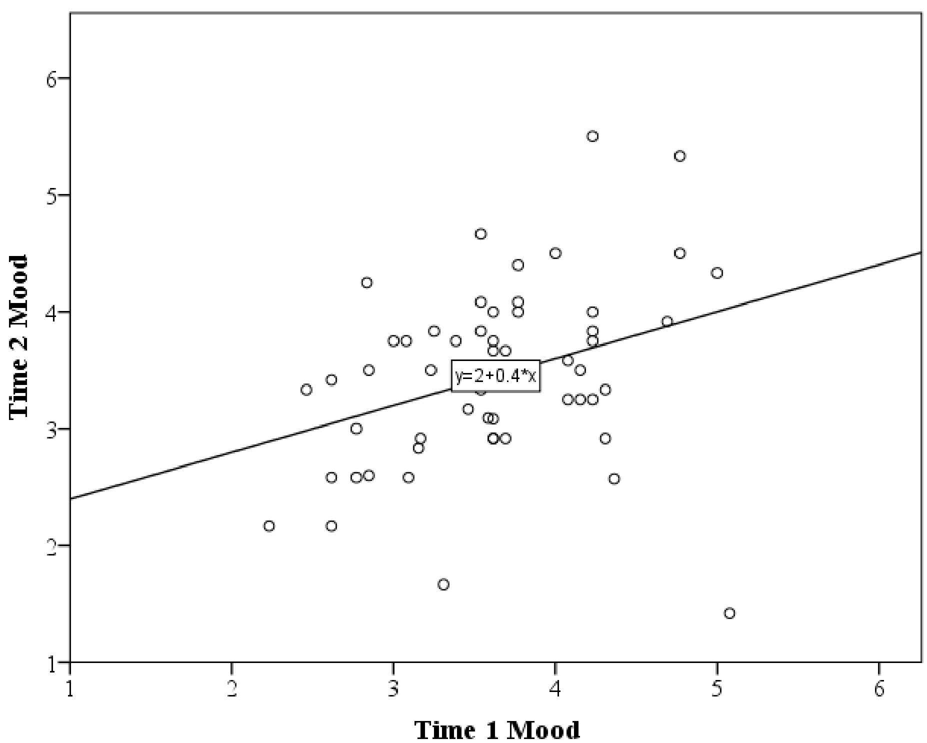
Figure 4

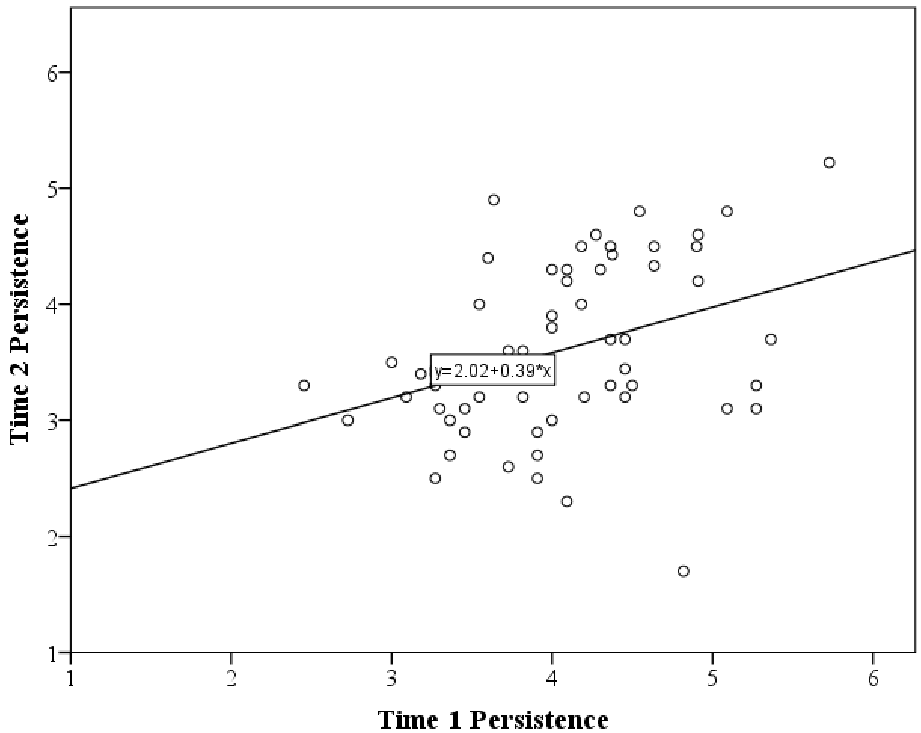
Predicting Time 2 Temperament from Time 1 Temperament (All Baby Siblings)






y=2.14+0.35*x
Time 2 Approach
Time 1 Approach
1
2
3
4
5
6


y=2+0.4*x
Time 2 Mood
Time 1 Mood
1
2
3
4
5
6


y=2.02+0.39*x
Time 2 Persistence
Time 1 Persistence
1
2
3
4
5
6

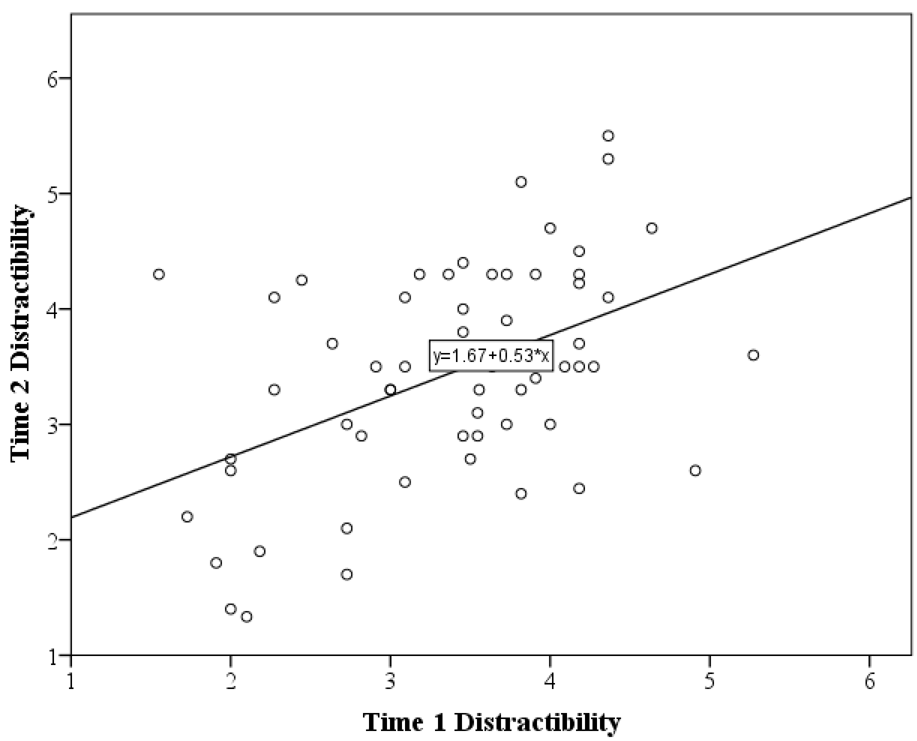

y=1.67+0.53*x
Time 2 Distractibility
Time 1 Distractibility
1
2
3
4
5
6

Figure 5

Simple Slopes for Time 2 Temperament on Time 1 Temperament, at levels of Time Lag

Key

- - - Low (-1 SD) Time Lag — Average Time Lag – – – High (+1 SD) Time Lag

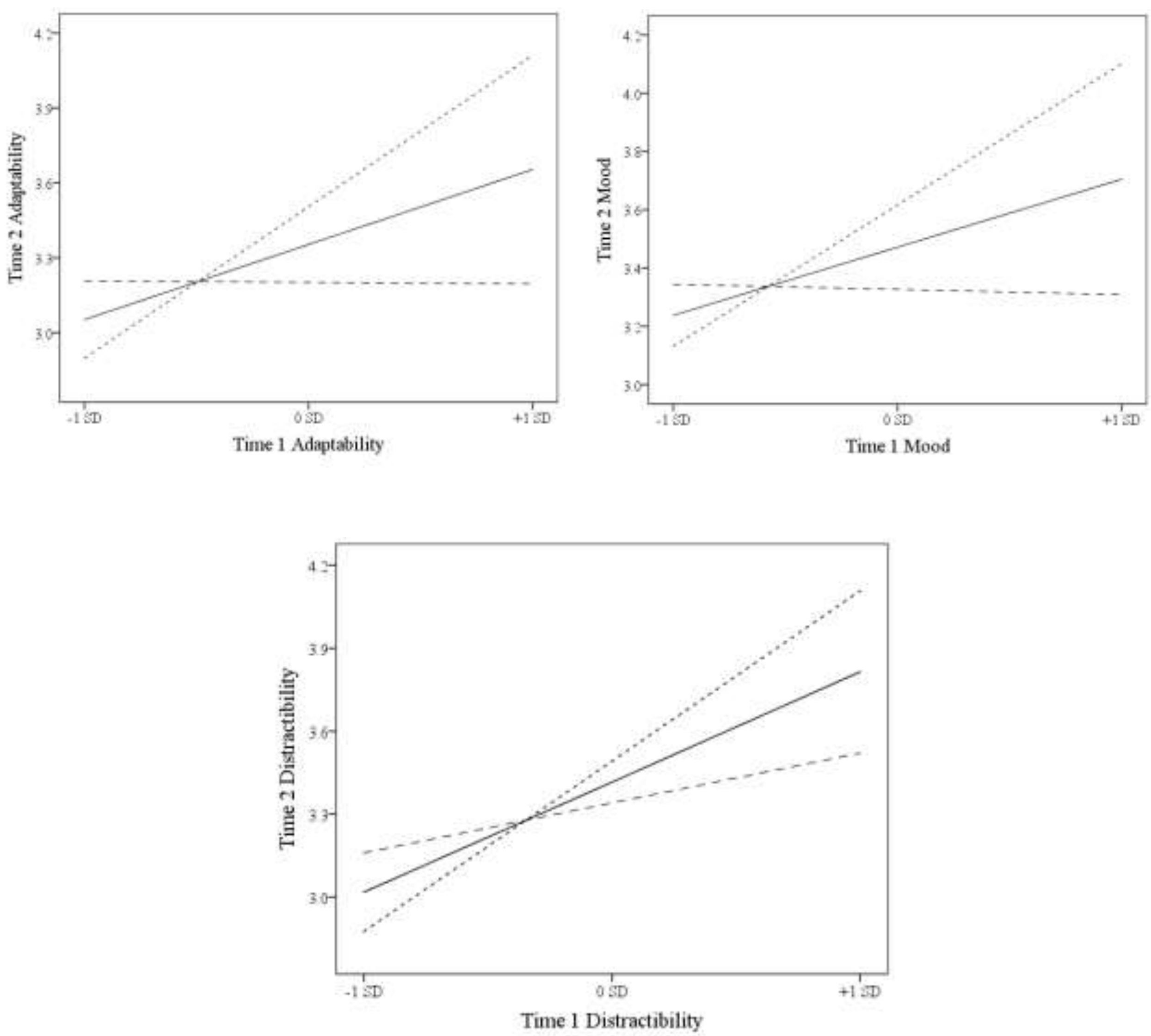
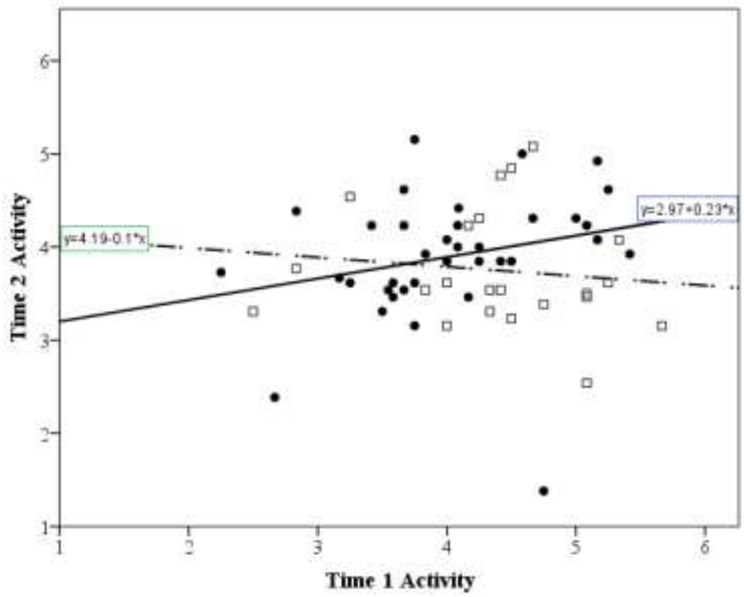


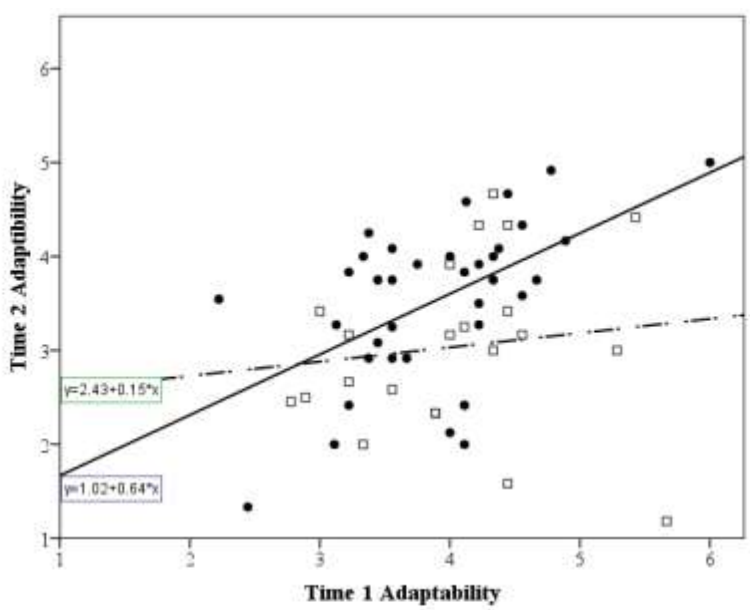
Figure 6

Predicting Time 2 Temperament from Time 1 Temperament (By Time 1 Diagnostic Group)

Key

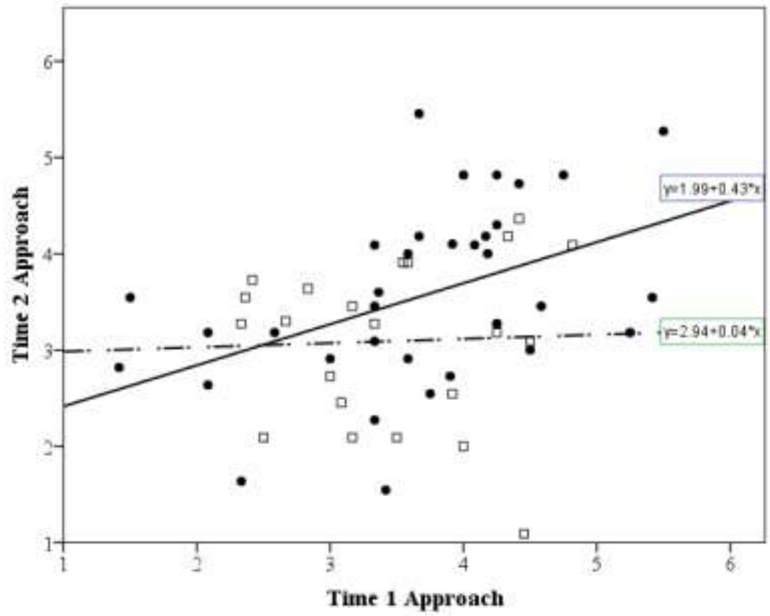
● ASD □ non-ASD — ASD ▪ – ▪ non-ASD

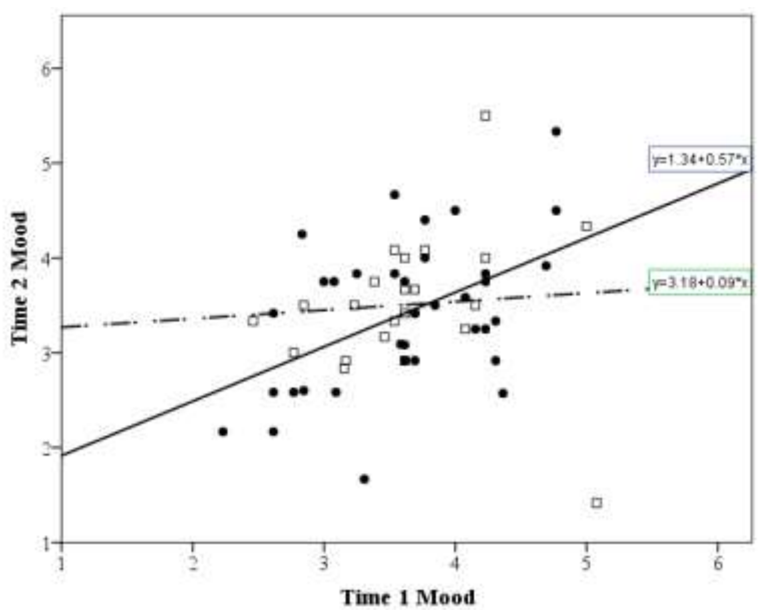




Key

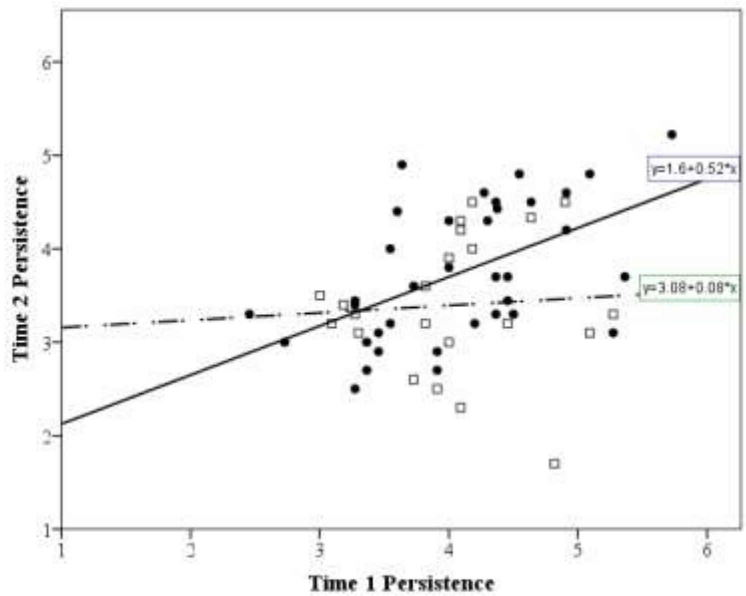
● ASD □ non-ASD —— ASD ▪ — ▪ non-ASD





Key

● ASD □ non-ASD —— ASD ▪ — ▪ non-ASD



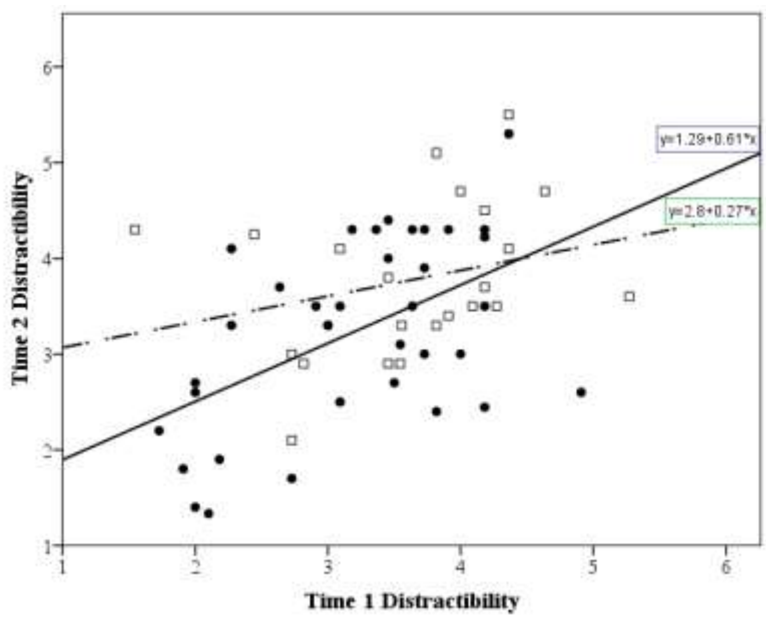
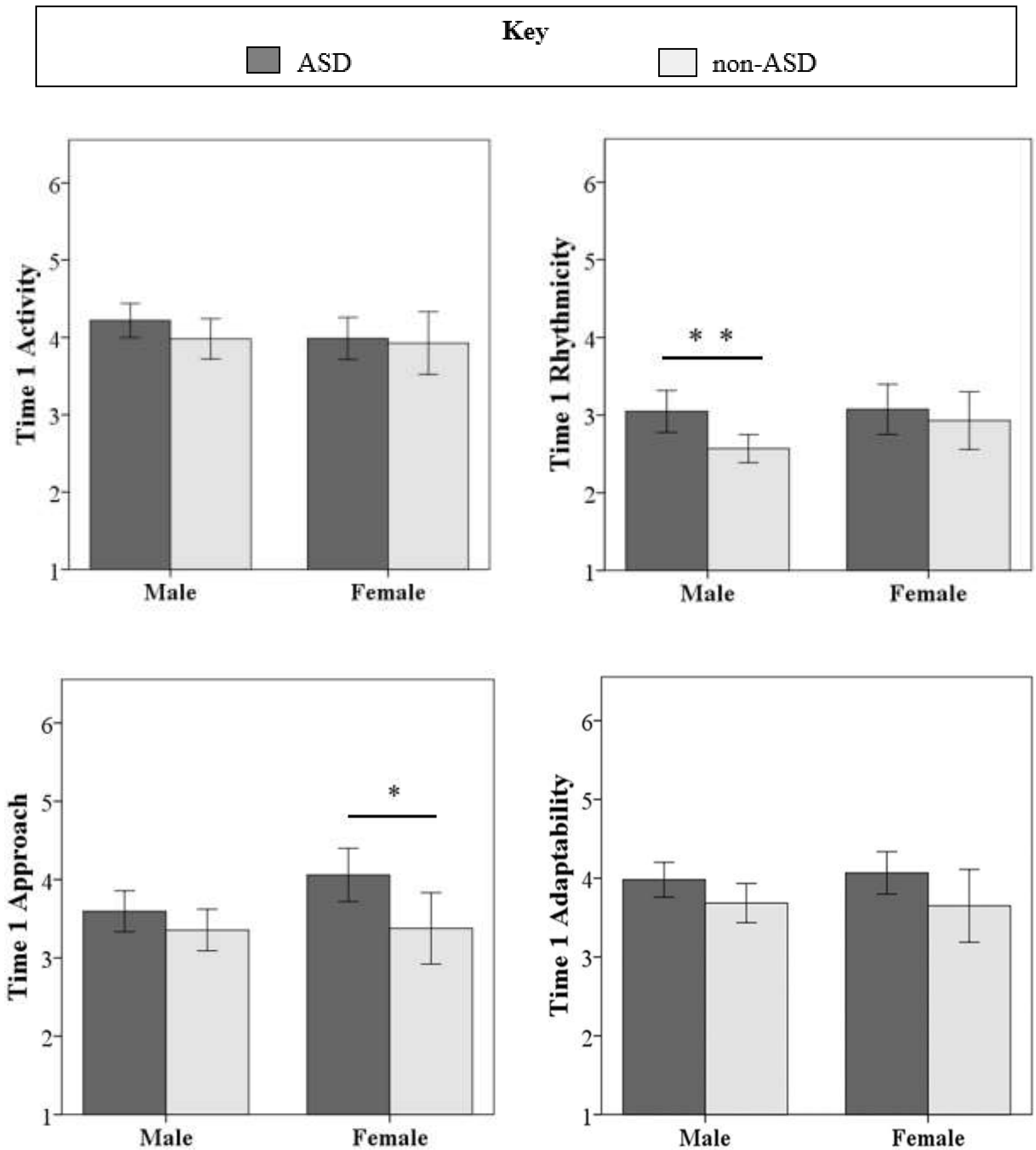
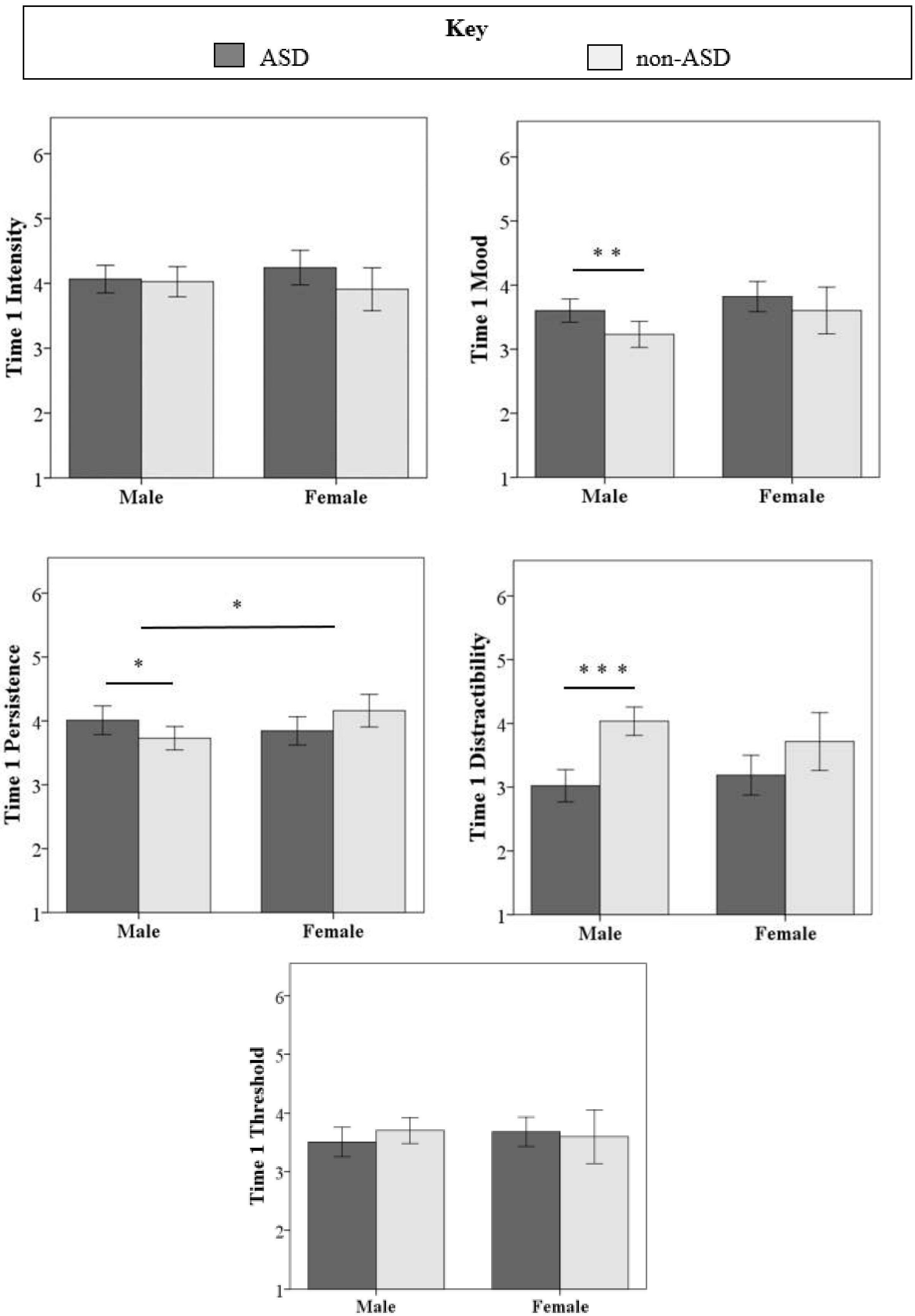


Figure 7

Time 1 Temperament by Diagnostic Group and Gender



Note: Error bars: ±2 SE; * $p < .05$, ** $p < .01$, *** $p < .001$


Key
ASD
non-ASD
Time 1 Intensity
6
5
4
3
2
1
Male
Female
Time 1 Mood
6
5
4
3
2
1
* *
Male
Female
Time 1 Persistence
6
5
4
3
2
1
*
*
Male
Female
Time 1 Distractibility
6
5
4
3
2
1
* * *
Male
Female
Time 1 Threshold
6
5
4
3
2
1
Male
Female