UNIVERSITY OF CALIFORNIA SAN DIEGO

Social Attention and Mirroring Faces: Utilizing Eye Tracking and EEG Mu Suppression toward Biomarkers for Autism Spectrum Disorder

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Cognitive Science

by

Adrienne Moore

Committee in charge:

Professor Jaime Pineda, Chair
Professor Karen Pierce, Co-Chair
Professor Andrea Chiba
Professor Eric Courchesne
Professor Sarah Creel

2018

The Dissertation of Adrienne Moore is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California San Diego

2018

DEDICATION

This dissertation is dedicated to my father, Calvin M. Moore (1940 – 2004) for his unconditional support (and for keeping the bookshelves full of science fiction novels and National Geographic magazines); and to my mother, Holly Kilpatrick, for her decades of maternal behavior (and for keeping the bookshelves full of classical literature, plus agreeing that a liberal arts bachelor's degree was a good idea for me).

EPIGRAPH

“The mutual dependence of men is so great in all societies that scarce any human action is entirely complete in itself, or is performed without some reference to the actions of others…

…inference and reasoning concerning the actions of others enters so much into human life that no man, while awake, is ever a moment without employing it.”

-- David Hume (1777). Enquiries Concerning Human Understanding, Section VIII, Part 1.

We are so used to empathy that we take it for granted; yet it is essential for human society as we know it.

Instead of evolution having replaced simpler forms of empathy with more advanced ones, the latter are merely elaborations on the former and remain dependent on them. This also means that empathy comes naturally to us. It is not something we only learn later in life, or that is culturally constructed.

-- Frans de Waal (2005). The Compassionate Instinct, Part I. The Science of Human Goodness; “The Evolution of Empathy”.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge many people at UC San Diego without whom this accomplishment would not have been possible for me. First, I'd like to thank the Pineda Lab, especially Jaime Pineda for first introducing me to a range of captivating topics in Cognitive Neuroscience, and for kindly welcoming me back to the Cognitive Science Department after a lengthy hiatus, and for his wonderful ability to make neuroscience topics fascinating. Second, I'd like to thank the ACE Lab, especially Karen Pierce and Eric Courchesne, who have given me the chance to be deeply immersed in the fascinating world of autism research. Their commitment and contribution to innovative research and improving the lives of those with ASD is inspiring. Also, I'd like to thank the Temporal Dynamics of Learning Center, especially Andrea Chiba, for faithfully supporting me as a scientist, which kept me afloat through years of learning. And finally, I'd like to thank the Language Acquisition and Sound Recognition Lab, especially Sarah Creel, who introduced me to eye tracking and gave me a place of belonging in her lab.

Chapter 3, in full, is a reprint of the material as it appears in Molecular Autism, 2018. Moore, Adrienne; Wozniak, Madeline; Yousef, Andrew; Barnes, Cindy Carter; Cha, Debra; Courchesne, Eric; Pierce, Karen. Adrienne Moore was the primary author and investigator of this paper.

Chapter 4, in full, is a reprint of the material as it appears in Behavioural Brain Research, 2012. Moore, Adrienne; Gorodnitsky, Irina; Pineda, Jaime. Adrienne Moore was the primary author and investigator of this paper.

VITA

2018 - Doctor of Philosophy, University of California San Diego

2008 - Master of Science, University of California San Diego

2001 - Bachelor of Arts, St. John's College

Academic Employment

Project Manager (2013-2018)

UCSD Autism Center, Neurosciences Department

Center Co-Directors: Dr. Karen Pierce and Dr. Eric Courchesne

Laboratory Coordinator (2012-2013)

UCSD Language Acquisition and Sound Recognition Lab,

Cognitive Science Department

Lab Director: Dr. Sarah Creel

Student Advocate (2010)

UCSD, Summer Training Academy for Research in the Sciences

Program Director: Veronica Henson-Phillips

Teaching Assistant (2010)

UCSD, Cognitive Science Department

Lecturer: Marguerite McQuire; Course: Language/Communicative Disorders

Teaching Assistant (2009)

UCSD, Cognitive Science Department

Professor: Dr. Terry Jernigan; Course: Cognitive Neuroscience

Teaching Assistant (2008)

UCSD, Cognitive Science Department

Lecturer: Dr. Silvia Paparello; Course: Introduction to Cognitive Science

Teaching Assistant (2007)

UCSD, Cognitive Science Department

Professor: Dr. Jaime Pineda; Course: Neuropsychological Basis of Alternate States

Teaching Assistant (2005, 2006, 2010)

UCSD, Cognitive Science Department

Lecturer: Dr. Mary Boyle; Course: Minds and Brains

Publications

(in prep) Identifying prognostic biomarkers for children with autism spectrum disorder. Bacon, E.C., **Moore, A.,** Lee, Q., Carter Barnes, C., Pierce, K.

(in prep) Kwon, M.K., **Moore, A.,** Barnes, C. C., Cha, D., Courchesne, E., Pierce, K. Typical levels of eye-region fixation in toddlers with ASD across multiple contexts.

(in prep) Lombardo, M., Eyler, L., **Moore, A.,** Barnes, C. C., Cha, D., Courchesne, E., Pierce, K. Early heterogeneity and functional disconnection between default mode, visual, and attention networks in toddlers with autism.

Moore, A., Wozniak, M., Yousef, A., Barnes, C. C., Cha, D., Courchesne, E., & Pierce, K. (2018). The geometric preference subtype in ASD: identifying a consistent, early-emerging phenomenon through eye tracking. *Molecular Autism, 9*, 19.

(in prep) **Moore, A.**, & Ruvolo, P. (co-first-authors), O'Rorke, M., Wu, T., Chiba, A., Littlewort, G., Movellan, J., Bartlett, M. Facial mimicry responses predict the ability to detect deceit.

(in prep) **Moore, A**., Chiba, A., Paulus, M. Cardiorespiratory responses to viewing others in pain: emotional contagion and cold pressor distress.

Pineda, J., Bai, J.M., Aragon, O., Pelton, H., **Moore, A.** (2012). The Impact of plasticity-induced rehabilitation training on face processing in high functioning autism: a mu rhythm perspective. *Psychology of Social Cognition*, Nova Science Publishers.

Moore, A., Gorodnitsky, I., Pineda, J. (2012). EEG mu component responses to viewing emotional faces. *Behavioural Brain Research. 226* (1), 309-316.

Pineda, J., **Moore, A.R.,** Elfenbein, H., Cox, R. (2009). Hierarchically organized mirroring processes in social cognition: the functional neuroanatomy of empathy. *Mirror Neuron Systems,* Humana Press.

Conference Presentations

International Meeting for Autism Research (IMFAR) (2016). A new "geopref test" with complex social stimuli. (poster, first author and presenter)

Cognitive Neuroscience Society (CNS) Annual Meeting (2011). SOBI analysis of EEG mu mirroring responses to emotional faces. (poster, first author and presenter)

Temporal Dynamics of Learning Center (TDLC) All-Hands Meeting (2011). SOBI analysis of EEG mu mirroring responses to emotional faces. (poster, first author and presenter)

TDLC Social Interactions Network Meeting (2010). Mirroring of pain faces and deceit detection. (presentation, co-authored with Paul Ruvolo)

CNS Annual Meeting (2009). Empathy and simulation of the physiological response to pleasant and unpleasant thermal sensations. (poster, first author and presenter)

IMFAR (2009). A mu rhythm view of neurofeedback training effects of face processing impairments in autism. (poster, presented by co-author J.M. Bai)

CNS Annual Meeting (2007). Affective face processing and empathy, effects on the EEG mu rhythm. (poster, first author and presenter)

CNS Annual Meeting (2006). EEG investigation of mu rhythm suppression while viewing pictures of emotional facial expressions. (poster, presented by co-author S. Thurman)

Awarded

Temporal Dynamics of Learning Center (TDLC) Trainee Small Grant (2011)

TDLC San Diego Fellowship (2010)

Frieda Daum Urey Endowed Fellowship (2009)

TDLC Trainee Small Grant (2009)

TDLC San Diego Fellowship (2009)

Glushko Merit Award (2008)

TDLC Trainee Small Grant (2008)

TDLC San Diego Fellowship (2007)

APA Diversity Program in Neuroscience Predoctoral Fellowship (2006)

ABSTRACT OF THE DISSERTATION

Social Attention and Mirroring Faces: Utilizing Eye Tracking and EEG Mu Suppression toward Biomarkers for Autism Spectrum Disorder

by

Adrienne Moore

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2018

Professor Jaime Pineda, Chair
Professor Karen Pierce, Co-Chair

Autism spectrum disorder (ASD) is defined behaviorally, by persistent difficulties with social communication, plus restricted, repetitive behaviors; its neural bases are not fully understood. ASD is also highly heterogeneous, in some cases a mild condition, but in others severely disabling. Currently there are no biomarkers used to identify young children with ASD, or to stratify them as more or less severe. Rather, reliable diagnosis at a young age requires observation by extensively trained clinician specialists. This dissertation is in part an attempt to contribute to the search for neural and behavioral valid markers of ASD using eye tracking and EEG.

Studies 1 and 2 examine test performance on eye tracking tasks that pair social images and geometric images as an objective indicator of ASD in 12-48 months infants/toddlers. Study 1 examines sex differences in social attention and whether eye tracking during presentation of competing geometric and social images is equally effective for detecting ASD in female and male

toddlers. Increased social attention could serve to protect females from reaching diagnostic thresholds for ASD, impacting female rates of diagnosis. The hypothesis that high risk females (non-ASD but with a diagnosed older sibling) exhibit increased social attention is also explored. Study 2 examines the stability of eye tracking test performance as an ASD indicator when images of biological motion (kids dancing) are replaced with more complex social interactions and a range of positive and negative emotional expressions. Symptom severity differences between ASD subtypes defined by eye tracking results are shown. Combining multiple such eye tracking tests is proposed to improve sensitivity while maintaining high specificity. Study 3 considers EEG mu suppression reflecting mirroring (recruiting first person sensorimotor representations for simulating others during social perception) as a potential neural biomarker for ASD. This study of neurotypical adults was the first to show mirroring of emotional faces reflected in mu suppression. How this finding might lead to a mu suppression neural biomarker for detecting ASD at the individual subject level is discussed also. Collectively, these studies explore potential autism biomarkers that might one day be used to efficiently and objectively identify and stratify toddlers with ASD.

INTRODUCTION

Recognizing that others have a subjective world of first person experiences and feelings like oneself has a developmental trajectory. Piaget, for example, claimed that babies are solipsists and that knowledge of other minds develops across the first year of life (Piaget, 1951). Some precursors may be present from birth, such as a perceptual bias towards attending to people and faces (Johnson, 2005). There are early manifestations of development of the insight of other minds as young infants spontaneously mimic faces, mapping the observed behavior of others to the sensory feedback from being in their own skin (Rayson, Bonaiuto, Ferrari, & Murray, 2017). Ultimately, the adult brain, in the course of daily life, almost continuously makes assumptions about what others are thinking and feeling. These assumptions appear to be supported in part by mirroring, cortical simulation processes akin to the early mimicry and sensory feedback mentioned above (Gallese, Keysers, & Rizzolatti, 2004). These mirroring processes recruit neural systems for first person actions and sensations in order to understand others (Iacoboni, 2009). It's remarkable to think that these processes are so effective that by adulthood, the thoughts and feelings of others are as real to us as objects of perception, although they are entirely filled in or inferred as we have no direct knowledge of or access to any mental content but our own.

One can imagine how disruption to these pervasive social inference processes we generally take for granted would profoundly interfere with typical human behavior. Indeed, many of the core characteristics of autism spectrum disorder (ASD) are disruptions in this social competence (Tanguay, 2011). While some self-advocates argue persuasively that ASD in high functioning individuals is a valid neurological variation that should be understood and accommodated rather than treated medically (Armstrong, 2015; Kapp, Gillespie-Lynch, Sherman, & Hutman, 2013), around a third of people with ASD also have intellectual disability (Van Naarden Braun et al., 2015); 83% have at least one co-occurring developmental diagnosis, and 16% have a co-occurring neurological diagnosis (Levy et al., 2010). Early,

intensive behavioral interventions can have a profound impact on many with ASD, and the earlier the age of starting treatment, the larger the gains predicted (Wallace & Rogers, 2010). This motivates the search for signs of ASD during the infant and early toddler stages. There is evidence from post-mortem tissue for prenatal brain differences in ASD, so there ought to be signals of ASD that can be detected even during the first year of life (Courchesne et al., 2018). However, the early processes that are disrupted in ASD are not fully understood, and ASD identification often doesn't occur until age three or four in the US, and later in many other parts of the world (Durkin et al., 2015). The current best practices for ASD diagnosis rely on clinician judgments of observed behaviors, are not fully objective, and do not utilize the growing body of neuroscientific research on ASD (Hus, 2017; Lord et al., 2012; Randall et al., 2018).

Further, age of ASD identification for females lags behind that of males, and the efficacy of many screening and diagnostic tools for females with ASD is not fully tested (Van Wijngaarden-Cremers et al., 2014). Behavioral research suggests that sex differences in toddler communication and social behavior in ASD and non-ASD toddlers alike confer an advantage for females, which may influence their susceptibility to ASD (Barbu, Cabanes, & Le Maner-Idrissi, 2011). However, the large sex difference in ASD prevalence (approximately 4:1 males:females) may also be impacted by ascertainment biases due to common ASD clinical tools being optimized to detect males with autism (Werling, 2016). For all these reasons, identification of robust, objective indicators or biomarkers of ASD has great potential value for the future of ASD research and clinical practices (Voineagu & Yoo, 2013).

CHAPTER 1

SOCIAL ATTENTION AND MIRRORING IN AUTISM SPECTRUM DISORER AND NEUROTYPICAL BRAIN AND BEHAVIOR

This dissertation addresses social attention and mirroring, proceeding from development to adulthood, both in people who are neurotypical and atypical, and studied at the levels of behavior and neural activity. Eye tracking provides behavioral data on tasks very similar to those used by cognitive neuroscience (e.g. viewing videos of social images under controlled conditions), and so occupies an intermediate level linking a considerable body of information on neural systems with the observations of complex behaviors made by clinicians (Falck-Ytter, Bolte, & Gredeback, 2013). Still, the National Institutes of Mental Health's number one strategic objective is to define the mechanisms of complex behaviors in biological terms, from molecular to cellular to brain networks (Insel, 2009). Therefore, the EEG research portion of this dissertation adds a vital brain systems component by investigating neural activity reflecting social information processing, a domain known to function abnormally in ASD.

Organization of the Dissertation

Part One will concern detecting toddlers with ASD, in the hopes of maximizing treatment impact through early intervention. This will involve using eye tracking and the brain's early emerging bias toward viewing people and social scenes.

Chapter 2 will leverage a very large sample of GeoPref test data to ask whether this eye tracking test effectively identifies females, and whether the same relationships between clinical measures and eye tracking performance are found in females as in males. The GeoPref test is a stimulus video that simultaneously presents abstract geometric motion and human biological motion (kids moving energetically) as separate video streams side by side. When combined with eye tracking, the GeoPref test identifies a subset of children with autism, on the basis of increased

time spent looking at the geometric images, which over a certain threshold is unique to toddlers who go on the receive an ASD diagnosis. As early phenotypic presentation of ASD differs in males and females in ways that have only been recently and partially explored, a sample large enough to contain a sizeable group of females with ASD (n=911 subjects, 266 ASD, 59 females with ASD) will be examined. This large sample also allows the testing of a hypothesis related to sex differences between male and female baby siblings who have an older sibling with ASD but are not themselves ASD. That is, this chapter will examine whether at-risk but non-ASD females exhibit enhanced social attention compared to at-risk non-ASD males, which may serve as a protective factor against development of ASD.

Chapter 3, which is already published (Moore et al., 2018), reports a second eye tracking test, the Complex Social test, which alters the social images presented in the original GeoPref test while holding the geometric stimuli constant. Specifically, the Complex Social test differs from the original by having increased depictions of emotional expressions and longer, more complicated scenes including dyadic interactions. This chapter shows that the Complex Social test identifies a subset of ASD toddlers from a screened, general population cohort by their preference for geometric compared to social stimuli, and that the clinical profile of these ASD toddlers includes greater ADOS symptom severity. Further, while the original GeoPref test's specificity is very high, its sensitivity is fairly low. Therefore, Chapter 3 examines whether the combined use of the original GeoPref and Complex Social tests improves this low sensitivity.

Part Two will address brain processes involved in understanding other minds, specifically while viewing emotional facial expressions. More specifically, it will discuss and assess the simulating social brain of adults reflected in EEG mu rhythm suppression. If mu suppression proves to be a reliable index of mirroring at the individual subject level, it could become a tool with great clinical utility for ASD.

Chapter 4, which is already published (Moore, Gorodnitsky, & Pineda, 2012), describes application of a blind source separation (BSS) algorithm, SOBI (Second Order Blind Identification)

(Tang, Sutherland, & McKinney, 2005), to show that event-related desynchronizations (ERDs) occur in sensorimotor mu EEG components during an emotional face viewing task. While evidence of similar mirroring activity had previously been published for viewing hand actions, mu suppression in response to viewing emotional facial expressions had not previously been investigated. This study also shows that facial expressions of positive and negative valence perturb the mu rhythm differently, at multiple post-stimulus latencies in the right hemisphere only.

Chapter 5 will have two parts. First, it will review some of the literature on mu suppression and mirroring that has emerged since the 2012 publication of Chapter 4. The focus will be on papers suggesting best methodological practices for measuring mu suppression as an index of mirroring, as this may be critical for utilization of mu suppression as a clinical tool for ASD. Second, Chapter 5 will be briefly revisited in light of recent literature on mu suppression, and a few additional experimental results that were not included in Chapter 4 will be discussed as well.

Chapter 6 will be a brief, overarching conclusion to the dissertation that summarizes findings from Chapters 2, 3, and 4, and discusses translation of ASD research findings to create public health impact.

Hypotheses

Chapter 2:

- The GeoPref test for ASD identification is predicted to work equally well for early detection of both males and females with ASD, i.e. with sensitivity around 20% and specificity around 98%.

- Eye tracking scores are predicted to have the same relationships to clinical phenotype measures in females as in males, i.e. female and male ASD toddlers who prefer geometric images will have more severe impairment than ASD toddlers who prefer social images.

- Considering a set of unaffected younger siblings to an older sibling with ASD, females are predicted to show enhanced social attention on the GeoPref test compared to males.

Chapter 3:

- A new Complex Social stimulus is predicted to improve test performance for detection of toddlers with ASD over the original GeoPref test by increasing the frequency and duration of dyadic interactions and emotional faces depicted during eye tracking.

- Combined use of the original and Complex Social GeoPref tests is predicted to improve test performance for detection of toddlers with ASD, i.e. to increase the sensitivity.

Chapter 4:

- Mirroring of emotional facial expressions is predicted to be measureable with EEG mu component suppression, with components extracted based on SOBI blind source separation.

- Perception of positive and negative faces are predicted to impact mu component suppression differently.

- Individual differences in mu suppression when viewing emotional faces are predicted to correlate with individual differences in trait empathy (tested in Chapter 4 but discussed in Chapter 5).

Definitions of Key Terms

ERD, ERS: Event-related desynchronization, event-related synchronization. Decrease or increase respectively in the synchronization of neuronal populations that give rise to an EEG signal, time-locked to an event, measured as decrease or increase respectively in power in a particular frequency band relative to a baseline.

Mu suppression: Reduction in power in the mu band (approximately 8-13 Hz) EEG signal from sensorimotor cortex, measured at central electrodes sites (C3 and C4). When it occurs in the absence of actual movement, but during perception of hands or faces, it is often interpreted as an indicator of cortical mirroring or simulation of another person's actions.

Mirroring: Functions that allow us to process the actions and emotions of others by replicating or simulating them, internally (in the case of mirroring in the cortex) or externally (in the case of mirroring with the body, e.g. mimicry and emotional contagion). This does not need to be mediated by “mirror neurons” per se.

Facial mimicry: To respond with congruent facial expressions when viewing an emotional face. This usually refers to an automatic, unconscious reaction.

Emotional contagion: The tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person and, consequently, to converge emotionally.

GeoPref test: An eye tracking paradigm that pairs abstract, geometric images with social images and is used to identify certain infants and toddlers with ASD with notably high specificity through preferential viewing of geometric images.

Sensitivity: An indicator for a diagnostic test defined as the number that will be correctly identified as positive by the test, out of all the people who are truly positive for the tested disorder.

Specificity: An indicator for a diagnostic test defined as the number that will be correctly identified as negative by the test, out of all the people who are truly negative for the tested disorder.

PPV: Positive predictive value. An indicator for a diagnostic test defined as the number of correct, true positives (actual instances of the disorder), out of all the test results that are positive.

NPV: negative predictive value. An indicator for a diagnostic test defined as the number of correct, true negatives (actually not instances of the disorder), out of all the test results that are negative.

Conceptual Background

Social attention and mirroring in typical development

Two sets of classical psychological studies of social behavior were conducted with newborns in maternity wards within hours of birth. Mark Johnson established that newborns have an innate preference for and tendency to orient toward faces (Johnson, Dziurawiec, Ellis, & Morton, 1991). The eye tracking tasks that are the focus of Part One (Moore et al., 2018; Pierce et al., 2016) utilize this infant tendency. This innate bias toward person perception from birth helps specialize the social brain, fine-tuning regions to expertly process faces, emotions, and eye gaze (Johnson, 2005; Tomalski, Csibra, & Johnson, 2009). However, infants don't just preferentially perceive faces, they also mimic them (Meltzoff, 2007; Vincini, Jhang, Buder, & Gallagher, 2017b), which relates to Part Two's focus on mirroring. Andrew Meltzoff's classical studies of newborns asserted that within hours of birth infants mimic various mouth gestures (Meltzoff & Moore, 1977, 1983). Recent work with rigorous controls and better experimental designs has cast doubt on the claim that facial mimicry is immediately present in neonates (Oostenbroek et al., 2016; Vincini, Jhang, Buder, & Gallagher, 2017a). Still, research has consistently found facial mimicry is present very early, well within the first year of life, by 5 and 7 months (Isomura & Nakano, 2016; Kaiser, Crespo-Llado, Turati, & Geangu, 2017) for specific forms of mimicry. Recent studies have also reported pupillary dilation in response to viewing dilated pupils in 4 to 6 month infants as a new form of early facial mimicry (Fawcett, Arslan, Falck-Ytter, Roeyers, & Gredeback, 2017; Fawcett, Wesevich, & Gredeback, 2016).

Preferential perception of people and automatic mimicry of their faces occur widely in adults as well, and lead to emotional contagion, or adopting aspects of an observed emotion (Dimberg & Thunberg, 2012; Hess & Blairy, 2001; Prochazkova & Kret, 2017). This emotional contagion occurs not just for faces and basic emotions, but for crying in infants (Geangu, Benga,

Stahl, & Striano, 2010), gestures and postures (Chartrand & Lakin, 2013), yawns (Norscia & Palagi, 2011), blinks and pupil size (Kret & De Dreu, 2017), and skin conductance and heart rate change (Balconi & Maria Elide Vanutelli, 2017), and appears to occur across several species (Davila Ross, Menzler, & Zimmermann, 2008; Demuru & Palagi, 2012; Palagi, Nicotra, & Cordoni, 2015). Thus mimicry, plus experiencing one's own contagiously acquired corresponding feeling state, and attributing the introspectively perceived feeling back to the perceived other, might yield the notion of other minds with subjective states (Meltzoff & Moore, 1997). Perhaps these processes of vicarious brain and bodily activity for the actions, emotions and sensations of others bring infants out of solipsism, by making the connection that the external behavior of others implies internal states and other minds (Meltzoff, 2007).

This infant self-other learning process proceeds bi-directionally during social interactions, as not only do infants mimic their parents, parents frequently mimic their infants as well, showing them the external behavior that matches the internal state the child is experiencing at that moment (Murray et al., 2016; Rayson et al., 2017). Infant predilection for orienting to caregiver faces and gestures ensures that they have opportunities to create these foundational self-other mappings during the first year of life (Frank, Vul, & Johnson, 2009). Later, toddlers deliberately imitate others, gesturally, vocally and when acting on objects, which broadens their early self-other mappings, and scaffolds language learning and mature theory of mind (S. S. Jones, 2009; Young et al., 2011). However, the emergence of mirroring occurs prior to goal directed action understanding, deliberate imitation and theory of mind, (Singer, 2006; Suddendorf, Oostenbroek, Nielsen, & Slaughter, 2013; Tousignant, Eugene, & Jackson, 2017), and many researchers consider the neural bases of mirroring (automatic, preconceptual mapping of self and other) and mentalizing (thinking about what others are thinking) to be distinct processes subserved by distinct brain networks (Adolphs, 2009; Coricelli, 2005; Pineda & Hecht, 2009).

Research on mirroring in adults

The mirroring version of this model stresses that mimicry and emotional contagion don't necessarily have to actually activate the body peripherally, but rather simulation mechanisms in the social brain networks for empathy and mirroring facilitate understanding what others are feeling and doing (Adolphs, 2006). In other words, perceived human actions (e.g. smiling) are mapped cross-modally: that is, action representations (e.g. motoric execution of smiling) and perceptual representations of the first person consequences of actions (e.g. proprioceptive, interoceptive, emotional, and somatosensory perception of what it feels like to smile) are associated with and activated during the visual perception of another's action (e.g. smiling done by someone else).

As evidence that this embodied simulation is causally related to processing the facial emotions of others, working backward from a group of 108 neuropsychological patients with various focal lesions, Ralph Adolph found that the lesions that most disrupt the ability to recognize facial emotion expressions were those to somatosensory cortex, especially in the right hemisphere (Adolphs, Damasio, Tranel, Cooper, & Damasio, 2000). Relatedly, several studies of mu suppression argue that somatosensory cortex is the most likely source of the mu signal (Arnstein, Cui, Keysers, Maurits, & Gazzola, 2011; Pfurtscheller & Lopes da Silva, 1999; Ritter, Moosmann, & Villringer, 2009), consistent with mu suppression induced by viewing faces of different emotions which is reported in Chapter 4. While much early research on mirroring focused on premotor regions and action representations (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), other approaches take a wider view of mirroring as subserved by an extended mirroring system that also includes sensorimotor cortex and the insula (Gallese et al., 2004; Pineda, 2008) or by multiple simulating social brain networks for both mirroring actions and empathy for emotions and experienced bodily states (Kennedy & Adolphs, 2012).

Many fMRI studies of social perceptual tasks also support the conclusion that visual

and auditory perception of others activates premotor, somatosensory, proprioceptive and interoceptive representations of ourselves for a wide range of tasks. Perception of point-light biological motion (compared to scrambled point light displays) significantly activates premotor cortex (Saygin, Wilson, Hagler, Bates, & Sereno, 2004). Both experiencing disgust (induced by inhaling noxious odorants) and observing someone else's reaction of disgust activate the same anterior insula sites underlying interoceptive visceral perception (Gazzola et al., 2012) (Wicker et al., 2003). Observing pain, both facial expressions of pain and hands and feet in painful positions, recruits many of the same cortical representations activated when we experience pain ourselves, including premotor, somatosensory, anterior cingulate and anterior insula regions (Jackson, Meltzoff, & Decety, 2005) (Budell, Jackson, & Rainville, 2010). "Ourselves" in this context usually means in our bodies, and somatotopic organization distinguishing hands and faces in sensorimotor cortices are often apparent in imaging studies of social simulations and mirroring (Ehrsson, Geyer, & Naito, 2003; Leslie, Johnson-Frey, & Grafton, 2004; Stippich, Ochmann, & Sartor, 2002). Several studies have also shown common activation for pleasant and unpleasant first person and vicarious touch in primary and secondary somatosensory cortex and the insula (Keysers et al., 2004) (Gazzola et al., 2012) (Lamm, Silani, & Singer, 2015). Both perceiving and imitating basic emotional facial expressions activate many of the same key areas noted above (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003), especially in the right hemisphere (Leslie et al., 2004). Furthermore, activity level in somatosensory cortex, intraparietal sulcus and premotor cortex during a facial imitation task has been found to be correlated with a trait empathy score (the Empathy Quotient) (Braadbaart, de Grauw, Perrett, Waiter, & Williams, 2014).

There has been widespread agreement across decades that the mu rhythm is related to activity somewhere around the central sulcus, which divides primary somatosensory from primary motor cortex, with cortical sources in "sensorimotor cortex" (Arroyo et al., 1993; Nam,

Jeon, Kim, Lee, & Park, 2011; Pineda, 2005). More detailed localization is perhaps important for understanding the functional significance of the mu rhythm, and how it corresponds to the fMRI studies of mirroring described above. While motor and premotor activity probably contribute somewhat to the signal (Arroyo et al., 1993; Pineda, 2005), somatosensory cortex seems to be the primary source of the mu rhythm, based on multiple types of evidence including MEG and combined EEG/fMRI studies. The mu rhythm has peaks in power around both 10 Hz and 20 Hz, however, this dissertation has focused on the classic alpha mu rhythm around 10 Hz only. Earlier EEG research viewed the 20 Hz peak as merely a harmonic of the 10 Hz peak (Pfurtscheller & Lopes da Silva, 1999), while more recent reports show they are functionally distinct. MEG has been used to investigate the mu rhythm in mirroring responses during observation of hand actions and observation of faces, both dynamic faces and still photos (Hari & Salmelin, 2012) (Nishitani & Hari, 2002). MEG research consistently finds the signals from ~10 and ~20 Hz mu to have different generators (Hari, 2006) and traces the ~20 Hz mu back to the precentral motor cortex, and the ~10 Hz mu to postcentral somatosensory cortex (Caetano, Jousmaki, & Hari, 2007; Salmelin, Hamalainen, Kajola, & Hari, 1995). At least three additional studies (e.g. combined EEG and fMRI studies) draw the same conclusion (Arnstein et al., 2011; Coll, Press, Hobson, Catmur, & Bird, 2017; Ritter et al., 2009). (Ritter et al., 2009) used blind source separation to identify the mu rhythm on data from an fMRI manual motor task, and found ~10 Hz activity inversely related to post-central (somatosensory) and ~20 Hz to pre-central (motor) cortices. (Arnstein et al., 2011) also report that mu suppression (at 10Hz) covaries with the BOLD signal during execution and observation of hand actions in primary somatosensory cortex, as well as inferior parietal lobe (which is consistently reported in mirroring network research), and dorsal premotor cortex.

Social attention and mirroring in ASD

Many points along the process sketched out above, of early emerging perceptual bias

and bidirectional mimicry linked to emotional contagion and cortical mirroring simulations, measured with fMRI or EEG, have been reported to function differently in people with ASD. For facial mimicry, findings are mixed, with some studies reporting deficits and others reporting typical behavior. For example, deficits in rapid and spontaneous but not voluntary mimicry of facial expressions found using EMG were reported for children 8-12, adolescents, and adults with ASD (Clark, Winkielman, & McIntosh, 2008; McIntosh, Reichmann-Decker, Winkielman, & Wilbarger, 2006; Oberman, Winkielman, & Ramachandran, 2009). However, a similar study of 6 and 7 year olds found normal EMG in the ASD group (Deschamps, Coppes, Kenemans, Schutter, & Matthys, 2015), as did a second study in children and adolescents with ASD (Schulte-Ruther et al., 2017). For the most part these studies had few subjects, and the subjects were high-functioning, older children or adults, and the developmental studies only included typically developing contrast groups, not children with non-ASD delays.

One comprehensive study of toddlers with ASD compared to those with other developmental disorders did address orofacial imitation (Rogers, Hepburn, Stackhouse, & Wehner, 2003). With 24 children with autism (mean age 34 months), 18 children with fragile X syndrome, 20 with other developmental disorders (DD), and 15 typically-developing (TD) children, these investigators found significantly greater impairment for the ASD group for overall imitation and orofacial imitation compared to all other groups. For fragile X, imitation was strongly influenced by whether the child also met ASD criteria, and in ASD, imitation skills were strongly negatively correlated with ASD symptom severity and joint attention ability after controlling for developmental level. This study seems very promising, though overall facial mimicry in autism does not seem very thoroughly studied, and autism researchers have not fully adopted standard definitions or tasks (Sevlever & Gillis, 2010) (e.g. distinguishing mimicry from imitation and emotional contagion, or using EMG to measure mimicry).

For emotional contagion research, results related to ASD are also mixed and seem to

depend on the particular form of emotional contagion examined. For example, fear contagion was reportedly not triggered in people with ASD observing others expressing fear, but emotional contagion for pain was reported to be intact (Hadjikhani et al., 2009) (Hadjikhani et al., 2014). Contagious yawning was reportedly altered in ASD, but it was increased in PDD-NOS while decreased in autism (Helt, Eigsti, Snyder, & Fein, 2010). One study of toddlers with multiple contrast groups (26 with ASD (mean age 33 months), 24 DD, 15 TD) measured change in hedonic facial tone from videos (coded by raters blind to diagnosis) of interactions with the experimenter who displayed positive and negative valence emotion to arrive at an emotional contagion score (Scambler, Hepburn, Rutherford, Wehner, & Rogers, 2007). Interestingly, they found that emotional contagion responses occurred much less frequently (approximately half as often) in ASD than in comparison groups, and responses were also muted when they occurred. Further, correlations were reported between emotional contagion scores, joint attention measures, and ASD symptom severity. It would be interesting to see the additional results of autonomic measures of emotional contagion (e.g. heart rate changes, pupillary response, skin conductance) in a large developmental study of ASD and non-ASD toddlers, similar to these feasibility studies (Billeci et al., 2018; Di Palma et al., 2017), but this apparently has not been done thus far.

The mirroring literature in ASD related to cortical simulation of other people during perception is highly contentious (Gallese, Gernsbacher, Heyes, Hickok, & Iacoboni, 2011). For example, using static images of emotional facial expressions, (Dapretto et al., 2006) reported atypical fMRI mirroring activation in ASD children imitating and observing emotional faces; however, they focused on BA44 and BA45, which contain Broca's area and the homologs of area F5 where individual mirror neurons [1] were first found in macaques. This study was quickly

[1] Specific claims of mirror neurons in Brodmann Area 44 (the human homolog to macaque area F5) or the possible dysfunction of these neurons in ASD, is tangential to this dissertation and so

challenged, mainly on theoretical grounds (Southgate & Hamilton, 2008), but also as later failing to replicate (Gallese et al., 2011). However, the primary "replication" studies which did not show mirroring deficits in ASD used only hand images, not face stimuli (Williams et al., 2006) (Martineau, Andersson, Barthelemy, Cottier, & Destrieux, 2010). Most of the subsequent debate about mirror system deficits in ASD has revolved around observing hand actions and understanding actions or inferring goals and intentions (e.g. (Dinstein et al., 2010) (Hickok, 2009)). A subsequent fMRI study of viewing faces also found decreased mirroring activity in ASD in face-related somatosensory and premotor cortex, as well as in other face processing regions (e.g. FFA, STS and amygdala), and did not emphasize "mirror neuron" theory, and was not similarly challenged (Hadjikhani, Joseph, Snyder, & Tager-Flusberg, 2007). Furthermore, a systematic review article by a mirror neuron theory challenger, while stating there's little evidence for global mirror system dysfunction in ASD, and that studies of the mirror neuron system using observation of hand actions often fail to find ASD group differences, concluded mirroring studies using emotional stimuli do consistently report ASD group differences (Hamilton, 2013).

In addition to fMRI, the mirroring literature for ASD also contains many studies based on EEG mu suppression. This is especially relevant for developmental research, because EEG is more feasible than fMRI for studying young, developmentally delayed children who are awake (Marshall & Meltzoff, 2011). Mu suppression refers to reduction in power of an EEG signal from sensorimotor cortex (indicating activation of sensorimotor neurons), and when this occurs in the absence of actual movement, but during perception of social stimuli, it is often taken to indicate mirroring, or simulation of the other person's actions (Pineda, 2005). The EEG mirroring in ASD story overall is similar to that of the mirroring system and ASD fMRI research described above.

not covered here, but for references on testing for mirror neurons while recording from individual neurons in humans see (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010) and (Perry et al., 2018).

The first descriptions of possible ASD deficits in mirroring measured as EEG mu suppression (Bernier, Dawson, Webb, & Murias, 2007; Oberman et al., 2005) were received with excitement, and then criticism (Fan, Decety, Yang, Liu, & Cheng, 2010; Gallese et al., 2011). More than a dozen studies of mu suppression report differences in ASD compared to neurotypical participants for mirroring tasks (for review see (Hobson & Bishop, 2017)). However, there are notable arguments against the mirroring interpretation of much of the mu suppression data that differs between ASD and TD groups, related to baseline selection and the influence of alpha EEG on mu suppression studies (Hobson & Bishop, 2017; Hobson & Bishop, 2016), discussed more thoroughly in Chapters 4 and 5.

A fairly consistent, early-emerging finding, usually revealed through eye tracking experiments while children watch video stimuli, is reduced time attending to social stimuli in ASD and an atypical manner of viewing social information when it is attended (Chita-Tegmark, 2016). Eye tracking tasks show point light biological motion displays are not preferred to scrambled point light displays in toddlers and ASD children aged 3-7 years ((Annaz, Campbell, Coleman, Milne, & Swettenham, 2012). Reduced time looking at a social scene overall, and specifically at an actress within a scene, and more specifically at the actress's face, have been reported in ASD infants compared to high risk (due to having an ASD sibling) non-ASD infants and controls. This was reported as young as 6 months of age, in a prospective study that followed babies longitudinally until they received a diagnostic evaluation at age two (Chawarska, Macari, & Shic, 2013). Another prospective study of high risk 6 month olds, some of whom went on receive an ASD diagnosis at age two, found less time looking to the inner features of the face during speaking portions of a naturalistic video characterized the infants who went on to receive an ASD diagnosis (Shic, Macari, & Chawarska, 2014).

Further, the salience of competing items presented with social stimuli differentially modulates social attention in young ASD children compared to contrast groups (Kwon, Moore,

Pierce (2018) in press), whether or not the competing images relate to common restricted, repetitive ASD interests ((Sasson, Elison, Turner-Brown, Dichter, & Bodfish, 2011; Sasson, Turner-Brown, Holtzclaw, Lam, & Bodfish, 2008). Following the direction of another's gaze appears to be used less effectively in ASD toddlers, which is apparent by 11 to 18 months (Bedford et al., 2012), suggesting impaired knowledge of the referential nature of eye gaze (Brooks & Meltzoff, 2005; Guillon, Hadjikhani, Baduel, & Roge, 2014).

One area of social attention research with less consistent results is that of reduced time spent looking specifically to the eyes in infants and toddlers with ASD (Falck-Ytter & von Hofsten, 2011). One lab consistently finds decreased eye looking and increased mouth looking in ASD toddlers (W. Jones, Carr, & Klin, 2008), including differences during the first year of life. They report social orienting to eyes appears to be normal in the first months in ASD, but then declines, leading to significant differences by six months (W. Jones & Klin, 2013). However, we and other labs have not reproduced reduced eye looking or increased mouth looking in ASD infants and toddlers in similar studies (Kwon, Moore, Pierce (2018) in press) (Hosozawa, Tanaka, Shimizu, Nakano, & Kitazawa, 2012) (Chawarska, Macari, & Shic, 2012). This may be partly due to important dependencies between mouth looking and speech development in toddlers with and without language delays (Falck-Ytter et al., 2013). Infants with ASD also often fail to attend to child-directed speech ((Sperdin & Schaer, 2016), and the addition of speech to social scenes can amplify eye tracking differences between ASD and TD children (Chawarska et al., 2012).

Others have pointed out that visual attention in ASD is more circumscribed, perseverative and detail oriented across both social and nonsocial stimulus domains (Sasson et al., 2008). Some general visual processing differences not specifically tied to social stimuli are also observed in ASD children, including an increased number of lateral and downward gazes (Mottron, Dawson, Soulieres, Hubert, & Burack, 2006). These gaze behaviors might be

adaptive and function to limit hypersensitivity to visual stimuli by filtering high frequency signals by peripheral viewing (Noris et al., 2011). Slower visual attentional disengagement from a central stimulus is another impairment that presents itself during the first year in ASD development, and it is also not limited to processing social stimuli (Sacrey, Armstrong, Bryson, & Zwaigenbaum, 2014).

Sex difference in ASD visual social attention has recently begun to receive some attention in eye tracking research (Bishop, Veenstra-VanderWeele, & Sanders, 2016). Circumscribed interests in ASD were compared using eye tracking as children viewed arrays of objects, and results showed that ASD female attention, like ASD male attention, was more circumscribed and perseverative than that of controls (Harrop et al., 2018). However, ASD females preferred female-typical objects, and ASD males preferred male-typical objects, which was noted to be contrary to an "extreme male brain" theory of ASD that predicts male-typical viewing patterns in ASD females. Intriguingly, (Chawarska, Macari, Powell, DiNicola, & Shic, 2016) reported enhanced social attention (more time viewing social scenes and faces within the scene) in female infants with high familial risk (i.e. siblings to a proband with ASD), compared to high risk males and to low risk females. This is reminiscent of the "female protective effect", the finding that females require a greater genetic load of risk factors to yield an ASD diagnostic outcome than males due to some unidentified, sex-specific protective mechanism (Werling & Geschwind, 2013). This will be followed up in Chapter 2.

Although attentional processes, especially for social stimuli, are clearly disrupted in infancy in ASD, there are many closely related processes in early social development that also must function correctly in order for mature social cognition to emerge, and mirroring may be one of them. According to a recent conceptual framework for understanding ASD brain and behavior, it is disrupted sensorimotor and attentional processes that primarily contribute to the pre-symptomatic first year of ASD (Piven, Elison, & Zylka, 2018). However, the authors do not

explore the function of sensorimotor processes in social cognition, including activation of sensorimotor brain during perception of others (mirroring), although sufficient evidence suggests related disruptions in ASD (for a review see (Happe & Frith, 2014)). This is a relatively under explored domain in early autism research, and more attention should be given to sensorimotor simulations of others in early development of ASD.

Eye tracking based social attention measures have already been implemented as screening tools for pediatric office use in feasibility studies, and industry researchers have begun to produce validated multimodal biometric tools for ASD clinical trial usage that include eye tracking tests (Ness et al., 2017). A rigorously defined neural index of social brain activity that accurately and reliably distinguished ASD from other clinical populations would be a significant contribution to public health and NIMH research priorities (Kapur, Phillips, & Insel, 2012). NIH has embraced the promise of mirror neuron system research by funding a large program on the Functions and Development of the Mirroring Neuron System, establishing mu suppression as a future candidate biomarker for large-scale autism clinical trials as well (http://www.mirroringdevelopment.com/). Given the great toll of mental illnesses, developing new clinical measures to biologically ground the study and treatment of psychiatric and behavioral disorders is arguably one of the great social challenges of our age (Insel & Scolnick, 2006).

BIBLIOGRAPHY

Adolphs, R. (2006). How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Res, 1079*(1), 25-35. doi:10.1016/j.brainres.2005.12.127

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annu Rev Psychol, 60*, 693-716. doi:10.1146/annurev.psych.60.110707.163514

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J Neurosci, 20*(7), 2683-2690.

Annaz, D., Campbell, R., Coleman, M., Milne, E., & Swettenham, J. (2012). Young children with autism spectrum disorder do not preferentially attend to biological motion. *J Autism Dev Disord, 42*(3), 401-408. doi:10.1007/s10803-011-1256-3

Armstrong, T. (2015). The myth of the normal brain: embracing neurodiversity. *AMA J Ethics, 17*(4), 348-352. doi:10.1001/journalofethics.2015.17.4.msoc1-1504

Arnstein, D., Cui, F., Keysers, C., Maurits, N. M., & Gazzola, V. (2011). mu-suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal, and SI cortices. *J Neurosci, 31*(40), 14243-14249. doi:10.1523/jneurosci.0963-11.2011

Arroyo, S., Lesser, R. P., Gordon, B., Uematsu, S., Jackson, D., & Webber, R. (1993). Functional significance of the mu rhythm of human cortex: an electrophysiologic study with subdural electrodes. *Electroencephalogr Clin Neurophysiol, 87*(3), 76-87.

Balconi, M., & Maria Elide Vanutelli, M. E. (2017). Empathy in Negative and Positive Interpersonal Interactions. What is the Relationship Between Central (EEG, fNIRS) and Peripheral (Autonomic) Neurophysiological Responses? *Adv Cogn Psychol, 13*(1), 105-120. doi:10.5709/acp-0211-0

Barbu, S., Cabanes, G., & Le Maner-Idrissi, G. (2011). Boys and girls on the playground: sex differences in social development are not stable across early childhood. *PLoS One, 6*(1), e16407. doi:10.1371/journal.pone.0016407

Bedford, R., Elsabbagh, M., Gliga, T., Pickles, A., Senju, A., Charman, T., . . . team, B. (2012). Precursors to social and communication difficulties in infants at-risk for autism: gaze following and attentional engagement. *J Autism Dev Disord, 42*(10), 2208-2218. doi:10.1007/s10803-012-1450-y

Bernier, R., Dawson, G., Webb, S., & Murias, M. (2007). EEG mu rhythm and imitation impairments in individuals with autism spectrum disorder. *Brain Cogn, 64*(3), 228-237. doi:10.1016/j.bandc.2007.03.004

Billeci, L., Tonacci, A., Narzisi, A., Manigrasso, Z., Varanini, M., Fulceri, F., . . . Muratori, F. (2018). Heart Rate Variability During a Joint Attention Task in Toddlers With Autism Spectrum Disorders. *Front Physiol, 9*, 467. doi:10.3389/fphys.2018.00467

Bishop, S. L., Veenstra-VanderWeele, J., & Sanders, S. J. (2016). Attention Finally Being Paid to Girls at

Risk of Autism. *J Am Acad Child Adolesc Psychiatry, 55*(3), 159-160. doi:10.1016/j.jaac.2015.12.013

Braadbaart, L., de Grauw, H., Perrett, D. I., Waiter, G. D., & Williams, J. H. (2014). The shared neural basis of empathy and facial imitation accuracy. *Neuroimage, 84*, 367-375. doi:10.1016/j.neuroimage.2013.08.061

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Dev Sci, 8*(6), 535-543. doi:10.1111/j.1467-7687.2005.00445.x

Budell, L., Jackson, P., & Rainville, P. (2010). Brain responses to facial expressions of pain: emotional or motor mirroring? *Neuroimage, 53*(1), 355-363. doi:10.1016/j.neuroimage.2010.05.037

Caetano, G., Jousmaki, V., & Hari, R. (2007). Actor's and observer's primary motor cortices stabilize similarly after seen or heard motor actions. *Proc Natl Acad Sci U S A, 104*(21), 9058-9062. doi:10.1073/pnas.0702453104

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc Natl Acad Sci U S A, 100*(9), 5497-5502. doi:10.1073/pnas.0935845100

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annu Rev Psychol, 64*, 285-308. doi:10.1146/annurev-psych-113011-143754

Chawarska, K., Macari, S., Powell, K., DiNicola, L., & Shic, F. (2016). Enhanced Social Attention in Female Infant Siblings at Risk for Autism. *J Am Acad Child Adolesc Psychiatry, 55*(3), 188-195 e181. doi:10.1016/j.jaac.2015.11.016

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

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biol Psychiatry, 74*(3), 195-203. doi:10.1016/j.biopsych.2012.11.022

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Res Dev Disabil, 48*, 79-93. doi:10.1016/j.ridd.2015.10.011

Clark, T. F., Winkielman, P., & McIntosh, D. N. (2008). Autism and the extraction of emotion from briefly presented facial expressions: stumbling at the first step of empathy. *Emotion, 8*(6), 803-809. doi:10.1037/a0014124

Coll, M. P., Press, C., Hobson, H., Catmur, C., & Bird, G. (2017). Crossmodal Classification of Mu Rhythm Activity during Action Observation and Execution Suggests Specificity to Somatosensory Features of Actions. *J Neurosci, 37*(24), 5936-5947. doi:10.1523/JNEUROSCI.3393-16.2017

Coricelli, G. (2005). Two-levels of mental states attribution: from automaticity to voluntariness. *Neuropsychologia, 43*(2), 294-300. doi:10.1016/j.neuropsychologia.2004.11.015

Courchesne, E., Pramparo, T., Gazestani, V. H., Lombardo, M. V., Pierce, K., & Lewis, N. E. (2018). The ASD Living Biology: from cell proliferation to clinical phenotype. *Mol Psychiatry*. doi:10.1038/s41380-018-0056-y

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nat Neurosci, 9*(1), 28-30. doi:10.1038/nn1611

Davila Ross, M., Menzler, S., & Zimmermann, E. (2008). Rapid facial mimicry in orangutan play. *Biol Lett, 4*(1), 27-30. doi:10.1098/rsbl.2007.0535

Demuru, E., & Palagi, E. (2012). In bonobos yawn contagion is higher among kin and friends. *PLoS One, 7*(11), e49613. doi:10.1371/journal.pone.0049613

Deschamps, P. K., Coppes, L., Kenemans, J. L., Schutter, D. J., & Matthys, W. (2015). Electromyographic responses to emotional facial expressions in 6-7 year olds with autism spectrum disorders. *J Autism Dev Disord, 45*(2), 354-362. doi:10.1007/s10803-013-1890-z

Di Palma, S., Tonacci, A., Narzisi, A., Domenici, C., Pioggia, G., Muratori, F., . . . group, M. s. (2017). Monitoring of autonomic response to sociocognitive tasks during treatment in children with Autism Spectrum Disorders by wearable technologies: A feasibility study. *Comput Biol Med,* 85, 143-152. doi:10.1016/j.compbiomed.2016.04.001

Dimberg, U., & Thunberg, M. (2012). Empathy, emotional contagion, and rapid facial reactions to angry and happy facial expressions. *Psych J, 1*(2), 118-127. doi:10.1002/pchj.4

Dinstein, I., Thomas, C., Humphreys, K., Minshew, N., Behrmann, M., & Heeger, D. J. (2010). Normal movement selectivity in autism. *Neuron, 66*(3), 461-469. doi:10.1016/j.neuron.2010.03.034

Durkin, M. S., Elsabbagh, M., Barbaro, J., Gladstone, M., Happe, F., Hoekstra, R. A., . . . Shih, A. (2015). Autism screening and diagnosis in low resource settings: Challenges and opportunities to enhance research and services worldwide. *Autism Res, 8*(5), 473-476. doi:10.1002/aur.1575

Ehrsson, H. H., Geyer, S., & Naito, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. *J Neurophysiol, 90*(5), 3304-3316. doi:10.1152/jn.01113.2002

Falck-Ytter, T., Bolte, S., & Gredeback, G. (2013). Eye tracking in early autism research. *J Neurodev Disord, 5*(1), 28. doi:10.1186/1866-1955-5-28

Falck-Ytter, T., & von Hofsten, C. (2011). How special is social looking in ASD: a review. *Prog Brain Res, 189*, 209-222. doi:10.1016/B978-0-444-53884-0.00026-9

Fan, Y. T., Decety, J., Yang, C. Y., Liu, J. L., & Cheng, Y. (2010). Unbroken mirror neurons in autism spectrum disorders. *J Child Psychol Psychiatry, 51*(9), 981-988. doi:10.1111/j.1469-7610.2010.02269.x

Fawcett, C., Arslan, M., Falck-Ytter, T., Roeyers, H., & Gredeback, G. (2017). Human eyes with dilated pupils induce pupillary contagion in infants. *Sci Rep, 7*(1), 9601. doi:10.1038/s41598-017-08223-3

Fawcett, C., Wesevich, V., & Gredeback, G. (2016). Pupillary Contagion in Infancy: Evidence for Spontaneous Transfer of Arousal. *Psychol Sci, 27*(7), 997-1003. doi:10.1177/0956797616643924

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition, 110*(2), 160-170. doi:10.1016/j.cognition.2008.11.010

Gallese, V., Gernsbacher, M. A., Heyes, C., Hickok, G., & Iacoboni, M. (2011). Mirror Neuron Forum. *Perspect Psychol Sci, 6*(4), 369-407. doi:10.1177/1745691611413392

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn Sci, 8*(9), 396-403. doi:10.1016/j.tics.2004.07.002

Gazzola, V., Spezio, M. L., Etzel, J. A., Castelli, F., Adolphs, R., & Keysers, C. (2012). Primary somatosensory cortex discriminates affective significance in social touch. *Proc Natl Acad Sci U S A, 109*(25), E1657-1666. doi:10.1073/pnas.1113211109

Geangu, E., Benga, O., Stahl, D., & Striano, T. (2010). Contagious crying beyond the first days of life. *Infant Behav Dev, 33*(3), 279-288. doi:10.1016/j.infbeh.2010.03.004

Guillon, Q., Hadjikhani, N., Baduel, S., & Roge, B. (2014). Visual social attention in autism spectrum disorder: insights from eye tracking studies. *Neurosci Biobehav Rev, 42*, 279-297. doi:10.1016/j.neubiorev.2014.03.013

Hadjikhani, N., Joseph, R. M., Manoach, D. S., Naik, P., Snyder, J., Dominick, K., . . . de Gelder, B. (2009). Body expressions of emotion do not trigger fear contagion in autism spectrum disorder. *Soc Cogn Affect Neurosci, 4*(1), 70-78. doi:10.1093/scan/nsn038

Hadjikhani, N., Joseph, R. M., Snyder, J., & Tager-Flusberg, H. (2007). Abnormal activation of the social brain during face perception in autism. *Hum Brain Mapp, 28*(5), 441-449. doi:10.1002/hbm.20283

Hadjikhani, N., Zurcher, N. R., Rogier, O., Hippolyte, L., Lemonnier, E., Ruest, T., . . . Gillberg, C. (2014). Emotional contagion for pain is intact in autism spectrum disorders. *Transl Psychiatry, 4*, e343. doi:10.1038/tp.2013.113

Hamilton, A. F. (2013). Reflecting on the mirror neuron system in autism: a systematic review of current theories. *Dev Cogn Neurosci, 3*, 91-105. doi:10.1016/j.dcn.2012.09.008

Happe, F., & Frith, U. (2014). Annual research review: Towards a developmental neuroscience of atypical social cognition. *J Child Psychol Psychiatry, 55*(6), 553-557.

Hari, R. (2006). Action-perception connection and the cortical mu rhythm. *Prog Brain Res, 159*, 253-260. doi:10.1016/S0079-6123(06)59017-X

Hari, R., & Salmelin, R. (2012). Magnetoencephalography: From SQUIDs to neuroscience. Neuroimage 20th anniversary special edition. *Neuroimage, 61*(2), 386-396. doi:10.1016/j.neuroimage.2011.11.074

Harrop, C., Jones, D., Zheng, S., Nowell, S., Boyd, B. A., & Sasson, N. (2018). Circumscribed Interests and Attention in Autism: The Role of Biological Sex. *J Autism Dev Disord*. doi:10.1007/s10803-

018-3612-z

Helt, M. S., Eigsti, I. M., Snyder, P. J., & Fein, D. A. (2010). Contagious yawning in autistic and typical development. *Child Dev, 81*(5), 1620-1631. doi:10.1111/j.1467-8624.2010.01495.x

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *Int J Psychophysiol, 40*(2), 129-141.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J Cogn Neurosci, 21*(7), 1229-1243. doi:10.1162/jocn.2009.21189

Hobson, H. M., & Bishop, D. V. (2017). The interpretation of mu suppression as an index of mirror neuron activity: past, present and future. *R Soc Open Sci, 4*(3), 160662. doi:10.1098/rsos.160662

Hobson, H. M., & Bishop, D. V. M. (2016). Mu suppression - A good measure of the human mirror neuron system? *Cortex, 82*, 290-310. doi:10.1016/j.cortex.2016.03.019

Hosozawa, M., Tanaka, K., Shimizu, T., Nakano, T., & Kitazawa, S. (2012). How children with specific language impairment view social situations: an eye tracking study. *Pediatrics, 129*(6), e1453-1460. doi:10.1542/peds.2011-2278

Hus, Y. (2017). Issues in Identification and Assessment of Children with Autism and a Proposed Resource Toolkit for Speech-Language Pathologists. *Folia Phoniatr Logop, 69*(1-2), 27-37. doi:10.1159/000477398

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annu Rev Psychol, 60*, 653-670. doi:10.1146/annurev.psych.60.110707.163604

Insel, T. R. (2009). Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. *Arch Gen Psychiatry, 66*(2), 128-133. doi:10.1001/archgenpsychiatry.2008.540

Insel, T. R., & Scolnick, E. M. (2006). Cure therapeutics and strategic prevention: raising the bar for mental health research. *Mol Psychiatry, 11*(1), 11-17. doi:10.1038/sj.mp.4001777

Isomura, T., & Nakano, T. (2016). Automatic facial mimicry in response to dynamic emotional stimuli in five-month-old infants. *Proc Biol Sci, 283*(1844). doi:10.1098/rspb.2016.1948

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage, 24*(3), 771-779. doi:10.1016/j.neuroimage.2004.09.006

Johnson, M. H. (2005). Subcortical face processing. *Nat Rev Neurosci, 6*(10), 766-774. doi:10.1038/nrn1766

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*(1-2), 1-19.

Jones, S. S. (2009). The development of imitation in infancy. *Philos Trans R Soc Lond B Biol Sci, 364*(1528), 2325-2335. doi:10.1098/rstb.2009.0045

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch Gen Psychiatry, 65*(8), 946-954. doi:10.1001/archpsyc.65.8.946

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature, 504*(7480), 427-431. doi:10.1038/nature12715

Kaiser, J., Crespo-Llado, M. M., Turati, C., & Geangu, E. (2017). The development of spontaneous facial responses to others' emotions in infancy: An EMG study. *Sci Rep, 7*(1), 17500. doi:10.1038/s41598-017-17556-y

Kapp, S. K., Gillespie-Lynch, K., Sherman, L. E., & Hutman, T. (2013). Deficit, difference, or both? Autism and neurodiversity. *Dev Psychol, 49*(1), 59-71. doi:10.1037/a0028353

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry, 17*(12), 1174-1179. doi:10.1038/mp.2012.105

Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends Cogn Sci, 16*(11), 559-572. doi:10.1016/j.tics.2012.09.006

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron, 42*(2), 335-346.

Kret, M. E., & De Dreu, C. K. (2017). Pupil-mimicry conditions trust in partners: moderation by oxytocin and group membership. *Proc Biol Sci, 284*(1850). doi:10.1098/rspb.2016.2554

Lamm, C., Silani, G., & Singer, T. (2015). Distinct neural networks underlying empathy for pleasant and unpleasant touch. *Cortex, 70*, 79-89. doi:10.1016/j.cortex.2015.01.021

Leslie, K. R., Johnson-Frey, S. H., & Grafton, S. T. (2004). Functional imaging of face and hand imitation: towards a motor theory of empathy. *Neuroimage, 21*(2), 601-607. doi:10.1016/j.neuroimage.2003.09.038

Levy, S. E., Giarelli, E., Lee, L. C., Schieve, L. A., Kirby, R. S., Cunniff, C., . . . Rice, C. E. (2010). Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the United States. *J Dev Behav Pediatr, 31*(4), 267-275. doi:10.1097/DBP.0b013e3181d5d03b

Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D. M., . . . Risi, S. (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Arch Gen Psychiatry, 69*(3), 306-313. doi:10.1001/archgenpsychiatry.2011.148

Marshall, P. J., & Meltzoff, A. N. (2011). Neural mirroring systems: exploring the EEG mu rhythm in human infancy. *Dev Cogn Neurosci, 1*(2), 110-123. doi:10.1016/j.dcn.2010.09.001

Martineau, J., Andersson, F., Barthelemy, C., Cottier, J. P., & Destrieux, C. (2010). Atypical activation of the mirror neuron system during perception of hand motion in autism. *Brain Res, 1320*, 168-175. doi:10.1016/j.brainres.2010.01.035

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J. L. (2006). When the social

mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Dev Sci, 9*(3), 295-302. doi:10.1111/j.1467-7687.2006.00492.x

Meltzoff, A. N. (2007). The 'like me' framework for recognizing and becoming an intentional agent. *Acta Psychol (Amst), 124*(1), 26-43. doi:10.1016/j.actpsy.2006.09.005

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*(4312), 74-78.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Dev, 54*(3), 702-709.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining Facial Imitation: A Theoretical Model. *Early Dev Parent, 6*(3-4), 179-192. doi:10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R

Moore, A., Gorodnitsky, I., & Pineda, J. (2012). EEG mu component responses to viewing emotional faces. *Behav Brain Res, 226*(1), 309-316. doi:10.1016/j.bbr.2011.07.048

Moore, A., Wozniak, M., Yousef, A., Barnes, C. C., Cha, D., Courchesne, E., & Pierce, K. (2018). The geometric preference subtype in ASD: identifying a consistent, early-emerging phenomenon through eye tracking. *Mol Autism, 9*, 19. doi:10.1186/s13229-018-0202-z

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *J Autism Dev Disord, 36*(1), 27-43. doi:10.1007/s10803-005-0040-7

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Curr Biol, 20*(8), 750-756. doi:10.1016/j.cub.2010.02.045

Murray, L., De Pascalis, L., Bozicevic, L., Hawkins, L., Sclafani, V., & Ferrari, P. F. (2016). The functional architecture of mother-infant communication, and the development of infant social expressiveness in the first two months. *Sci Rep, 6*, 39019. doi:10.1038/srep39019

Nam, C. S., Jeon, Y., Kim, Y. J., Lee, I., & Park, K. (2011). Movement imagery-related lateralization of event-related (de)synchronization (ERD/ERS): motor-imagery duration effects. *Clin Neurophysiol, 122*(3), 567-577. doi:10.1016/j.clinph.2010.08.002

Ness, S. L., Manyakov, N. V., Bangerter, A., Lewin, D., Jagannatha, S., Boice, M., . . . Pandina, G. (2017). JAKE(R) Multimodal Data Capture System: Insights from an Observational Study of Autism Spectrum Disorder. *Front Neurosci, 11*, 517. doi:10.3389/fnins.2017.00517

Nishitani, N., & Hari, R. (2002). Viewing lip forms: cortical dynamics. *Neuron, 36*(6), 1211-1220.

Noris, B., Barker, M., Nadel, J., Hentsch, F., Ansermet, F., & Billard, A. (2011). Measuring gaze of children with Autism Spectrum Disorders in naturalistic interactions. *Conf Proc IEEE Eng Med Biol Soc, 2011*, 5356-5359. doi:10.1109/IEMBS.2011.6091325

Norscia, I., & Palagi, E. (2011). Yawn contagion and empathy in Homo sapiens. *PLoS One, 6*(12), e28472. doi:10.1371/journal.pone.0028472

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Brain Res Cogn Brain Res, 24*(2), 190-198. doi:10.1016/j.cogbrainres.2005.01.014

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2009). Slow echo: facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders. *Dev Sci, 12*(4), 510-520. doi:10.1111/j.1467-7687.2008.00796.x

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., . . . Slaughter, V. (2016). Comprehensive Longitudinal Study Challenges the Existence of Neonatal Imitation in Humans. *Curr Biol, 26*(10), 1334-1338. doi:10.1016/j.cub.2016.03.047

Palagi, E., Nicotra, V., & Cordoni, G. (2015). Rapid mimicry and emotional contagion in domestic dogs. *R Soc Open Sci, 2*(12), 150505. doi:10.1098/rsos.150505

Perry, A., Stiso, J., Chang, E. F., Lin, J. J., Parvizi, J., & Knight, R. T. (2018). Mirroring in the Human Brain: Deciphering the Spatial-Temporal Patterns of the Human Mirror Neuron System. *Cereb Cortex, 28*(3), 1039-1048. doi:10.1093/cercor/bhx013

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin Neurophysiol, 110*(11), 1842-1857.

Piaget, J. (1951). *Play, dreams, and imitation in childhood*. New York,: Norton.

Pierce, K., Marinero, S., Hazin, R., McKenna, B., Barnes, C. C., & Malige, A. (2016). Eye Tracking Reveals Abnormal Visual Preference for Geometric Images as an Early Biomarker of an Autism Spectrum Disorder Subtype Associated With Increased Symptom Severity. *Biol Psychiatry, 79*(8), 657-666. doi:10.1016/j.biopsych.2015.03.032

Pineda, J. A. (2005). The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing". *Brain Res Brain Res Rev, 50*(1), 57-68. doi:10.1016/j.brainresrev.2005.04.005

Pineda, J. A. (2008). Sensorimotor cortex as a critical component of an 'extended' mirror neuron system: Does it solve the development, correspondence, and control problems in mirroring? *Behav Brain Funct, 4*, 47. doi:10.1186/1744-9081-4-47

Pineda, J. A., & Hecht, E. (2009). Mirroring and mu rhythm involvement in social cognition: are there dissociable subcomponents of theory of mind? *Biol Psychol, 80*(3), 306-314.

Piven, J., Elison, J. T., & Zylka, M. J. (2018). Toward a conceptual framework for early brain and behavior development in autism. *Mol Psychiatry, 23*(1), 165. doi:10.1038/mp.2017.212

Prochazkova, E., & Kret, M. E. (2017). Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neurosci Biobehav Rev, 80*, 99-114. doi:10.1016/j.neubiorev.2017.05.013

Randall, M., Egberts, K. J., Samtani, A., Scholten, R. J., Hooft, L., Livingstone, N., . . . Williams, K. (2018). Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *Cochrane Database Syst Rev, 7*, CD009044. doi:10.1002/14651858.CD009044.pub2

Rayson, H., Bonaiuto, J. J., Ferrari, P. F., & Murray, L. (2017). Early maternal mirroring predicts infant

motor system activation during facial expression observation. *Sci Rep, 7*(1), 11738. doi:10.1038/s41598-017-12097-w

Ritter, P., Moosmann, M., & Villringer, A. (2009). Rolandic alpha and beta EEG rhythms' strengths are inversely related to fMRI-BOLD signal in primary somatosensory and motor cortex. *Hum Brain Mapp, 30*(4), 1168-1187. doi:10.1002/hbm.20585

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res, 3*(2), 131-141.

Rogers, S. J., Hepburn, S. L., Stackhouse, T., & Wehner, E. (2003). Imitation performance in toddlers with autism and those with other developmental disorders. *J Child Psychol Psychiatry, 44*(5), 763-781.

Sacrey, L. A., Armstrong, V. L., Bryson, S. E., & Zwaigenbaum, L. (2014). Impairments to visual disengagement in autism spectrum disorder: a review of experimental studies from infancy to adulthood. *Neurosci Biobehav Rev, 47*, 559-577. doi:10.1016/j.neubiorev.2014.10.011

Salmelin, R., Hamalainen, M., Kajola, M., & Hari, R. (1995). Functional segregation of movement-related rhythmic activity in the human brain. *Neuroimage, 2*(4), 237-243.

Sasson, N. J., Elison, J. T., Turner-Brown, L. M., Dichter, G. S., & Bodfish, J. W. (2011). Brief report: Circumscribed attention in young children with autism. *J Autism Dev Disord, 41*(2), 242-247. doi:10.1007/s10803-010-1038-3

Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S., & Bodfish, J. W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Res, 1*(1), 31-42. doi:10.1002/aur.4

Saygin, A. P., Wilson, S. M., Hagler, D. J., Jr., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *J Neurosci, 24*(27), 6181-6188. doi:10.1523/JNEUROSCI.0504-04.2004

Scambler, D. J., Hepburn, S., Rutherford, M. D., Wehner, E. A., & Rogers, S. J. (2007). Emotional responsivity in children with autism, children with other developmental disabilities, and children with typical development. *J Autism Dev Disord, 37*(3), 553-563. doi:10.1007/s10803- 006-0186-y

Schulte-Ruther, M., Otte, E., Adiguzel, K., Firk, C., Herpertz-Dahlmann, B., Koch, I., & Konrad, K. (2017). Intact mirror mechanisms for automatic facial emotions in children and adolescents with autism spectrum disorder. *Autism Res, 10*(2), 298-310. doi:10.1002/aur.1654

Sevlever, M., & Gillis, J. M. (2010). An examination of the state of imitation research in children with autism: Issues of definition and methodology. *Res Dev Disabil, 31*(5), 976-984. doi:10.1016/j.ridd.2010.04.014

Shic, F., Macari, S., & Chawarska, K. (2014). Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. *Biol Psychiatry, 75*(3), 231-237. doi:10.1016/j.biopsych.2013.07.009

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature

and implications for future research. *Neurosci Biobehav Rev, 30*(6), 855-863. doi:10.1016/j.neubiorev.2006.06.011

Southgate, V., & Hamilton, A. F. (2008). Unbroken mirrors: challenging a theory of Autism. *Trends Cogn Sci, 12*(6), 225-229. doi:10.1016/j.tics.2008.03.005

Sperdin, H. F., & Schaer, M. (2016). Aberrant Development of Speech Processing in Young Children with Autism: New Insights from Neuroimaging Biomarkers. *Front Neurosci, 10*, 393. doi:10.3389/fnins.2016.00393

Stippich, C., Ochmann, H., & Sartor, K. (2002). Somatotopic mapping of the human primary sensorimotor cortex during motor imagery and motor execution by functional magnetic resonance imaging. *Neurosci Lett, 331*(1), 50-54.

Suddendorf, T., Oostenbroek, J., Nielsen, M., & Slaughter, V. (2013). Is newborn imitation developmentally homologous to later social-cognitive skills? *Dev Psychobiol, 55*(1), 52-58. doi:10.1002/dev.21005

Tang, A. C., Sutherland, M. T., & McKinney, C. J. (2005). Validation of SOBI components from high-density EEG. *Neuroimage, 25*(2), 539-553. doi:10.1016/j.neuroimage.2004.11.027

Tanguay, P. E. (2011). Autism in DSM-5. *Am J Psychiatry, 168*(11), 1142-1144. doi:10.1176/appi.ajp.2011.11071024

Tomalski, P., Csibra, G., & Johnson, M. H. (2009). Rapid orienting toward face-like stimuli with gaze-relevant contrast information. *Perception, 38*(4), 569-578. doi:10.1068/p6137

Tousignant, B., Eugene, F., & Jackson, P. L. (2017). A developmental perspective on the neural bases of human empathy. *Infant Behav Dev, 48*(Pt A), 5-12. doi:10.1016/j.infbeh.2015.11.006

Van Naarden Braun, K., Christensen, D., Doernberg, N., Schieve, L., Rice, C., Wiggins, L., . . . Yeargin-Allsopp, M. (2015). Trends in the prevalence of autism spectrum disorder, cerebral palsy, hearing loss, intellectual disability, and vision impairment, metropolitan atlanta, 1991-2010. *PLoS One, 10*(4), e0124120. doi:10.1371/journal.pone.0124120

Van Wijngaarden-Cremers, P. J., van Eeten, E., Groen, W. B., Van Deurzen, P. A., Oosterling, I. J., & Van der Gaag, R. J. (2014). Gender and age differences in the core triad of impairments in autism spectrum disorders: a systematic review and meta-analysis. *J Autism Dev Disord, 44*(3), 627-635. doi:10.1007/s10803-013-1913-9

Vincini, S., Jhang, Y., Buder, E. H., & Gallagher, S. (2017a). Neonatal Imitation: Theory, Experimental Design, and Significance for the Field of Social Cognition. *Front Psychol, 8*, 1323. doi:10.3389/fpsyg.2017.01323

Vincini, S., Jhang, Y., Buder, E. H., & Gallagher, S. (2017b). An unsettled debate: Key empirical and theoretical questions are still open. *Behav Brain Sci, 40*, e401. doi:10.1017/S0140525X16001977

Voineagu, I., & Yoo, H. J. (2013). Current progress and challenges in the search for autism biomarkers. *Dis Markers, 35*(1), 55-65. doi:10.1155/2013/476276

Wallace, K. S., & Rogers, S. J. (2010). Intervening in infancy: implications for autism spectrum disorders. *J Child Psychol Psychiatry, 51*(12), 1300-1320. doi:10.1111/j.1469-7610.2010.02308.x

Werling, D. M. (2016). The role of sex-differential biology in risk for autism spectrum disorder. *Biol Sex Differ, 7*, 58. doi:10.1186/s13293-016-0112-8

Werling, D. M., & Geschwind, D. H. (2013). Understanding sex bias in autism spectrum disorder. *Proc Natl Acad Sci U S A, 110*(13), 4868-4869. doi:10.1073/pnas.1301602110

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664.

Williams, J. H., Waiter, G. D., Gilchrist, A., Perrett, D. I., Murray, A. D., & Whiten, A. (2006). Neural mechanisms of imitation and 'mirror neuron' functioning in autistic spectrum disorder. *Neuropsychologia, 44*(4), 610-621. doi:10.1016/j.neuropsychologia.2005.06.010

Young, G. S., Rogers, S. J., Hutman, T., Rozga, A., Sigman, M., & Ozonoff, S. (2011). Imitation from 12 to 24 months in autism and typical development: a longitudinal Rasch analysis. *Dev Psychol, 47*(6), 1565-1578. doi:10.1037/a0025418

CHAPTER 2

SEX DIFFERENCES IN SOCIAL ATTENTION IN TODDLERS WITH AND WITHOUT AUTISM SPECTRUM DISORDER

Abstract

Autism spectrum disorder (ASD) is diagnosed 3-4 times less frequently in females than in males, and subsequently the phenotype of females with ASD is less thoroughly researched. Females may present with subtle differences that can make ASD diagnosis more difficult, for example, a female advantage in the general population in processing social information may impact diagnosis of females with ASD. Here we explore sex differences in social attention using an eye tracking test (the GeoPref test) previously shown to be effective for detecting ASD in a predominantly male toddler sample. The GeoPref test presents competing social and geometric video stimuli while tracking toddler gaze, and in a 78% male ASD sample it has been shown that those toddlers who spend the majority of their time viewing the geometric images go on to receive an ASD diagnosis with specificity of 98% and sensitivity of 21%. Female specific validation statistics are calculated here. In addition, recent eye tracking research has found social attention in non-ASD but high risk (due to having an older sibling with ASD) infant females to be enhanced relative to males and to low risk females. Therefore this risk by sex interaction in the non-ASD sample is tested as well, with the prediction of increased time viewing GeoPref test social images in high risk females.

A sample of 911 toddlers from 11-48 months (mean 25 months), 266 with ASD (59 females with ASD), was used to examine sex differences. Eye tracking data from the GeoPref test plus ADOS, MSEL and VABS clinical scores were examined. Female ASD subgroups based on GeoPref test scores indicating high and low preference for viewing social images ("SocPref" and "GeoPref" subtypes, defined by >69% time viewing social and geometric

images respectively) were compared. In ASD and non-ASD groups, frequency of SocPref and GeoPref subtypes were compared by sex. In high risk non-ASD subgroups based on familial autism risk conveyed by having an older sibling with ASD, and low risk contrast toddlers, sex differences in GeoPref tests and clinical measures were examined also.

The independent sample replicated previous findings, as ASD children attended significantly more to geometric images than did all contrast groups (TD, DD and ASD features). No sex differences were found in TD or ASD groups on clinical measures (MSEL, ADOS, VABS) or in percent time viewing geometric vs social images. However, a greater proportion of the SocPref subtype was found in the non-ASD females group than in non-ASD males. The GeoPref test accurately classified females with ASD, with sensitivity of 20% and specificity of 99%. Comparing female ASD GeoPref and SocPref subgroups revealed a large difference in MSEL verbal scores between the groups. Comparing high and low risk non-ASD toddlers, a risk by sex interaction was found, confirming high risk females spent increased time attending to social images during eye tracking.

The GeoPref test, a measure of toddler reduced social attention, is comparably effective for detection of ASD in males (ROC AUC=.72) and females (AUC=.75), and did not yield significant sex differences in %Geo scores in any diagnostic group. Verbal development is a strength of ASD females with strong social attention during eye tracking, compared to females with poor social attention detected by the GeoPref test who are significantly more verbally impaired. Among non-ASD toddlers, more females strongly prefer viewing social images, and high risk non-ASD females spend elevated time viewing the GeoPref test's social images. Heterogeneity in ASD, including the possibility of subtype differences and sex differences, requires research such as this with large sample sizes.

Background

According to current estimates, among children diagnosed with autism spectrum disorder (ASD), there are four males for each female (Baio et al., 2018). Because they make up a smaller proportion of the ASD population, females with ASD are diagnosed and treated using paradigms developed from research focused primarily on ASD males (Van Wijngaarden-Cremers et al., 2014). Younger age at start of treatment (e.g. before 48 months) predicts larger gains during treatment for ASD (MacDonald, Parry-Cruwys, Dupere, & Ahearn, 2014; Vivanti, Dissanayake, & Victorian, 2016), however, age of ASD identification for females lags behind that of males (Begeer et al., 2013; Ratto et al., 2018; Shattuck et al., 2009). Therefore, understanding sex differences in ASD presentation in toddlers is important, for optimizing screening, diagnostic tools, and interventions towards the needs of ASD females as necessary (Lai, Baron-Cohen, & Buxbaum, 2015; Lai, Lombardo, Auyeung, Chakrabarti, & Baron-Cohen, 2015).

Clinical use of eye tracking technology may be the new frontier in ASD early intervention efforts, and a number of eye tracking based clinical tools are in development (Frazier et al., 2016; Fujioka et al., 2016). Clinical trial measures for indexing treatment responses, prognostic indicators, and tools to aid diagnosis, especially in regions where access to trained clinical psychologists is limited, are some of the proposed clinical uses for eye tracking in ASD (Murias et al., 2018; Vargas-Cuentas et al., 2017). As eye tracking tools are developed for clinical use in ASD, these should be validated for detecting both males and females, a task we undertake here. Further, the causes of the high ratio of males to females with an ASD diagnosis is a fundamental question for autism research that remains unanswered (Werling & Geschwind, 2013). In addition to concerns regarding clinical care, better understanding of females with and without ASD should help address this central, unresolved research question (Werling, 2016).

Phenotypic Sex Differences in ASD

The unidentified neurobiological mechanisms causing the greater prevalence of ASD in boys may also lead to sex differences in cognitive profiles, autism symptoms, and daily functional abilities. A number of past studies reported that females with ASD tend to have lower mean IQs and greater social/communicative impairment than males with ASD (Lord & Schopler, 1985; Tsai & Beisler, 1983). It has also been noted that the prevalences of females and males are less unequal among more severely impaired individuals (Volkmar, Szatmari, & Sparrow, 1993). The overall 4:1 sex difference in prevalence has been reported as high as 7-10:1 for individuals in the high functioning cognitive range (Frazier & Hardan, 2017).

However, ASD sex ratios and clinical characteristics of females with ASD may be changing, as the DSM undergoes various alterations, and as efforts to identify females, including high functioning females, with ASD increase (Halladay et al., 2015; Lai, Baron-Cohen, et al., 2015). A recent review and meta-analysis of research conducted after the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) was implemented argues that 3:1 is a better estimate of the male to female ratio in ASD than the often cited ratio of 4:1 (Loomes, Hull, & Mandy, 2017). And while some new studies continue to report sex differences in cognitive ability in ASD (Frazier, Georgiades, Bishop, & Hardan, 2014), a number of recent studies with toddlers have not found greater female impairment in verbal or nonverbal cognitive development, or in core social communicative deficits, in young children with ASD (Hull, Mandy, & Petrides, 2017).

For example, Messinger et al., using a large sample of over 1000 at-risk infants followed prospectively (including 59 girls with ASD) reported small sex differences in ASD cognitive ability (verbal and nonverbal) and in ASD symptom severity at 24 and 36 months, but all in the direction of females performing better than males (Messinger et al., 2015). They noted that sex by group interactions were absent, that is, sex differences found in ASD were

consistently those found in typical development as well (Constantino, 2016; Messinger et al., 2016). Similarly, Zwaigenbaum et al., found small sex differences in toddler autism presentation severity with a baby sibs cohort, in the direction of higher cognitive abilities and lower autism severity in females (Zwaigenbaum et al., 2012). However, among the cognitive indices only fine motor skills, but neither verbal nor nonverbal (visual reception) cognitive developmental scores, were different. Several additional studies reported no significant sex differences in ASD toddler cognitive abilities (Hartley & Sikora, 2009), or no differences in either cognitive ability or social-communicative symptom severity in male and female toddlers with ASD, as well as no group by sex interactions (Andersson, Gillberg, & Miniscalco, 2013; Reinhardt, Wetherby, Schatschneider, & Lord, 2015).

One sex difference that has been reported fairly consistently is fewer restricted, repetitive behaviors in females compared to males with ASD (Frazier et al., 2014; Mandy et al., 2012; Szatmari et al., 2012). However, this finding is reported to become consistently apparent for children with ASD over the age of six years (Harrop, Gulsrud, & Kasari, 2015; Van Wijngaarden-Cremers et al., 2014). Further, some studies suggest that restricted, repetitive behaviors in females differ in kind rather than in quantity or severity from those of males, with regard to circumscribed interests in particular (Harrop et al., 2018). Harrop, et al., investigated sex differences in ASD toddler play with respect to restricted behaviors and interests, and concluded that in females with ASD, circumscribed interests tend to be female-typical (e.g. play with dolls) and more often have social themes (Harrop, Green, Hudry, & Consortium, 2017). However, play complexity and amount of symbolic play (e.g. pretend play with dolls), while lower in ASD than TD toddlers, was largely equivalent in ASD females and ASD males (Harrop et al., 2017; Harrop et al., 2015). If ASD is under-identified in females without intellectual impairment, subtle differences in female presentation of restricted, repetitive behaviors and interests may contribute (Kreiser & White, 2014).

Social Attention Sex Differences in ASD

Eye tracking studies have shown reduction in attention to social stimuli to be a hallmark of ASD in toddlers (Chita-Tegmark, 2016), but little eye tracking research has examined ASD sex differences in toddler social attention. Typical newborns are reported to exhibit sex differences in social attention, in fact, as young as 36 hours in age males reportedly looked significantly longer at a mobile than females, and females looked significantly longer at faces than at the mechanical toy (Connellan, Baron-Cohen, Wheelwright, Batki, & Ahluwalia, 2000). In an eye tracking study of 3-8 month old infants, an age when children are believed to lack awareness of gender identity, females showed a large (Cohen's d >1.0) visual preference for a doll over a toy truck, while males compared to females showed a greater number of gaze fixations on the truck (d=.78) (Alexander, Wilcox, & Woods, 2009). However, social perceptual tasks that find sex differences in the typical population have not necessarily revealed parallel sex differences in ASD. For example, the "reading the mind in the eyes" facial emotion recognition task has repeatedly shown a small but significant neurotypical female advantage (Alaerts, Nackaerts, Meyns, Swinnen, & Wenderoth, 2011), and also has consistently shown impaired performance in adults and children with ASD (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Holt et al., 2014). However, in a study involving over 200 females with ASD, a clear absence of ASD sex differences was reported for this well validated task (Baron-Cohen et al., 2015).

Using eye tracking, Chawarska et al., (Chawarska, Macari, Powell, DiNicola, & Shic, 2016) examined sex differences in attending to a social vignette (an actress trying to engage the viewer with infant directed speech) in a prospective sample of 101 high risk infants (baby siblings with an older sibling with an ASD diagnosis) and 61 low risk controls at ages 6, 9, and 12 months. This report did not include categorical assignment of diagnostic status (ASD, non-ASD) to the infant siblings, and instead, taking a dimensional approach, contained a mixture

of ASD and non-ASD infants (Constantino, 2011). While there were no sex differences in the low risk group, they found a group by sex interaction, specifically, increased social attention in high risk female infants compared to high risk males, low risk males and low risk females. Further, greater social attention during the first year was significantly associated with autism symptoms (social affect scores) at age two. Enhanced social attention in high risk females warrants further exploration, as it may provide social learning opportunities early in development that could serve as a protective factor, mitigating the impact of ASD risk.

In the current study, we leverage a very large eye tracking data sample from toddlers with ASD and both typically developing and developmentally delayed non-ASD toddlers (n=911) from an eye tracking task simultaneously presenting dynamic social and geometric images, the GeoPref test (see Figure 1). Prior research has shown increased attention to geometric images in ASD toddlers viewing the GeoPref test, with a predominantly male sample (Pierce, Conant, Hazin, Stoner, & Desmond, 2011; Pierce et al., 2016). This effect is not just found at the level of group averages, in fact, individual subjects with ASD can be classified effectively (i.e. 98% specificity) by GeoPref test scores, as the ranges of scores for ASD and non-ASD toddlers are partly non-overlapping. These properties suggest the GeoPref test, at least for males, could function as a biomarker and aid in screening, diagnosis or clinical trials related to treatment response, and so must be tested robustly in females as well.

By combining a new, independent sample of n=469 subjects with the previously published sample, we address several hypotheses related to GeoPref test sex differences here. First, we predict that the GeoPref test is similarly effective as a tool for identifying both male and female toddlers with ASD. Therefore, we predict high specificity for identifying certain toddlers with ASD in both sexes, based on their viewing pattern of strong preference for abstract geometric rather than social images. Further, we predict that the subset of females with the strongest preference for geometric rather than social images (the “GeoPref subtype”) will also demonstrate elevated symptom severity in ADOS, MSEL and VABS scores compared

to the ASD female subset with the strongest preference for social images (the "SocPref subtype"), replicating what was reported previously, primarily for males (Pierce et al., 2016).

Human biological motion and abstract geometric motion compete for attention in the GeoPref test. The 62 second test video consists of 28 short scenes, 4 of which are depicted here. The primary test metric is the %Geo score (percent time viewing the geometric video divided by total time viewing either the social or geometric video). More than 1600 children ages 11-48 months have successfully completed the GeoPref test, both at UCSD and in pediatrician offices as a screening tool.

Figure 2.1: Example Scenes from the GeoPref Test Video

Second, we report sex differences from a battery of cognitive and autism specific standardized assessments (i.e. ADOS, MSEL and VABS) in toddlers with and without ASD. This study builds on existing literature by utilizing a sample based primarily on referral after general population screening rather than based on a baby sibs design. We also include developmentally delayed and ASD features contrast groups in addition to typically developing controls, and have an ASD female sample size (n=59) comparable to the largest studies to date of the female ASD toddler phenotype.

Third, we test the hypothesis put forward by Chawarska et al., that females with

elevated risk for ASD due to having an older sibling with ASD also show enhanced social attention (Chawarska et al., 2016). In the current study enhanced social attention is indexed as increased time spent viewing social images and decreased time viewing geometric images during eye tracking with the GeoPref test. We also differ from the prior study by using eye tracking data collected at a mean age of just under age two rather than age one, and our high and low risk samples have been assigned to categorical diagnostic groups and do not include ASD toddlers.

Methods

Participant Recruitment

From November 2008 through September 2015, 911 research participants between 11 and 48 months in age (mean 25.0 months) completed autism focused evaluations involving standardized clinical testing and an eye tracking based assessment (the GeoPref test). Data from 469 of these subjects have not been included in any prior manuscripts; the remainder were described in prior papers, however, not with respect to sex differences (Pierce, Conant, et al., 2011; Pierce et al., 2016). An additional 236 subjects (21%) attempted to complete the assessments and eye tracking session but were excluded from analysis for various reasons (e.g. tantrum during eye tracking).

The majority of the study sample were referred by their pediatrician due to concerns regarding warning signs of autism spectrum disorder or developmental delays. These pediatricians, as participants in the Get SET (Screening, Evaluation, Treatment) early detection program, screen all toddlers for developmental delays at 12, 18 and 24 months, using the Communication and Symbolic Behavior Scales (CSBS) Infant-Toddler Checklist parent questionnaire (Pierce, Carter, et al., 2011; Wetherby, Brosnan-Maddox, Peace, & Newton, 2008). Through this program, children who score in the range of concern (the lowest scoring

10%) are referred for prospective developmental evaluations at the University of California, San Diego's Autism Center, and then referred to appropriate early intervention treatment services if warranted. The remaining subjects either self-referred due to parental concern about their child's development, or participated as controls. Regardless of referral source, all participants completed the same sequence of assessments.

Diagnostic and Psychometric Assessments

All testing took place at the UCSD Autism Center. Diagnostic and psychometric assessments were completed with licensed, PhD level clinical psychologists. After entering the study based on screening and referral at 12, 18 and 24 months, subjects are followed longitudinally, most completing three visits, until receiving a final diagnosis between 30 and 48 months of age. When subjects completed eye tracking at more than one visit, the data included here are from the earliest available visit, and the clinical assessment data presented are from the same visit.

Participants were sorted into four diagnostic groups: autism spectrum disorder (ASD), ASD features, developmental delay (DD), and typical development (TD). ASD diagnosis was based on DSM-IV or DSM-V criteria probed with the Autism Diagnostic Observation Schedule (ADOS), module T, 1 or 2 as appropriate (Gotham, Risi, Pickles, & Lord, 2007; Luyster et al., 2009). The ASD features group had significant ASD symptoms and/or elevated ADOS scores during at least one evaluation but did not meet full DSM criteria for ASD at their final diagnostic evaluation. The Mullen Scales of Early Learning (MSEL) was used to assess cognitive development in areas of expressive language, receptive language, visual reception and fine motor abilities (Mullen, 1995). The Vineland Adaptive Behavior Scales (VABS) parent interview was used to assess functional skills, i.e. the ability to perform typical daily activities (Sparrow, 1984). The MSEL and VABS are the basis for assignment to the DD category, which includes children with transient or persistent language, fine motor or global developmental

delays or other concerns such as social emotional delay. The TD category includes children who tested within the typical range on ADOS, MSEL and VABS. The final overall sample were n=266 ASD (207 male, 59 female); n=52 ASD features (44 male, 8 female); n=309 DD (230 male, 79 female); n=284 TD (157 male, 127 females). In this sample, 480 (53%) were over 36 months at final diagnosis; 250 (27%) were 30-36 months; 62 (7%) were 24-30 months; 119 (13%) have not returned for their final planned visits and were assigned to a diagnostic category based on visits at ages 12-24 months.

Familial risk subgroups. To create non-ASD subgroups based on familial risk, the inclusion criterion for the high risk group was an older sibling with a medical or educational professional's confirmed ASD diagnosis. For the low risk group, the inclusion criterion was having an older sibling with no history of any developmental concerns according to parent report (if the older sibling did not complete our study), and having no other relative with ASD in the immediate family (parents, grandparents, siblings, half-siblings). If the older sibling was a past participant in our research, the older sibling was eligible if assessed by our psychologists as not meeting ASD criteria at any visit (i.e. siblings with ASD features and non-ASD delays were allowed if the sibling was evaluated by our lab and found not to meet criteria for ASD at any visit). The final sample was n=57 high risk (28 females, 29 males) and n=91 low risk (42 females, 49 males). The remainder of the non-ASD sample did not have an older sibling or the sibling's risk status could not be determined.

Stimulus Video and Eye Tracking Equipment and Procedures

The GeoPref test is a 62 second, silent video composed of adjacent rectangular areas of interest (AOIs) each 12.9 ^{o} by 9.1^{o}. One AOI, social, presents people outdoors (primarily children shown individually) energetically moving their arms, shoulders, heads and hips (during a yoga warm up). The other AOI, geometric, presents complex, abstract shapes in a variety of

colors, changing dynamically. The two AOIs are synchronized to move through 28 short scenes, switching simultaneously, but otherwise provide two unrelated streams of dynamic images.

Using corneal reflection techniques, a Tobii T-120 eye tracking system set to a 60 Hz sampling rate was used to record X-Y coordinates for the left and right eyes as participants sat 60-65 cm from a 17 inch computer monitor displaying the test video. Children were typically seated on a parent's lap (or occasionally in a car seat instead), and the parent was instructed to look at the ceiling or close her/his eyes and not to talk during testing. Five point infant calibration was performed prior to testing using Tobii Studio software, and eye tracking data were collected only if the calibration results fell within the parameters reported to yield an accuracy of 0.5^{o} in the manufacturer's published white papers. Data were excluded from analysis if the child attended to or was eye tracked for less than 30 seconds (approximately 50% of the test video), and for a variety of reasons detailed in the supplemental materials, primarily related to child behavior or parental noncompliance (e.g. interfering with the child's behavior). Data preprocessing consisted of applying a 35 pixel Tobii fixation filter to the data to identify gaze fixations, and then calculating total viewing time, and time viewing each AOI divided by total viewing time, that is, percent time viewing each AOI ("%Geo" and "%Soc").

Statistical Methods

Replicating clinical phenotype differences based on diagnostic group.

Independent sample (n=469). The independent sample refers to previously unpublished subjects used to replicate the results of prior papers on the GeoPref test. The scores reported for all toddlers from the ADOS are the social affect (SA) sub-score, the restricted, repetitive behavior (RRB) sub-score, and the total score. From the MSEL, scores reported are nonverbal developmental quotient (the mean of fine motor and visual reception

age equivalent sub-scores divided by chronological age) and the verbal developmental quotient (the mean of receptive language and expressive language age equivalent sub-scores divided by chronological age), or NV-DQ and V-DQ. From the VABS, the adaptive behavior composite score (ABC) is reported, which summarizes the socialization, motor, communication and daily living skills sub-scores. To confirm that scores for the ADOS, MSEL and VABS differed as predicted and as shown in prior research between ASD toddlers and the other three diagnostic groups, an ANOVA was performed for each dependent variable followed by Bonferroni-corrected pairwise comparisons of each diagnostic group to ASD. See Table 1 for these characteristics of the independent sample.

Replicating reduced social attention differences based on diagnostic group.

Independent sample (n=469). First, percentage of total looking time spent attending to any part of the approximately 1 minute long video between the four diagnostic groups (as an indication of general attentional and engagement differences) was assessed using a 1-way ANOVA with Bonferroni correction on total time viewing either AOI. Then, to confirm percentage of total looking time spent viewing the geometric AOI between the four diagnostic groups was elevated in ASD toddlers, indicating reduced social attention consistent with our findings reported in prior papers, an ANCOVA was performed with %Geo as the dependent variable, and age at eye tracking as a covariate. Partial eta-squared effect sizes are presented as well. A significant main effect was followed by Bonferroni-corrected t-tests (1 tailed) between ASD and other diagnostic groups. Usage of Bonferroni correction at several points in this manuscript is very conservative, but it replicates the prior manuscript (Pierce et al., 2016), and the differences are such that it does not impact findings.

Sex differences in clinical phenotype by diagnostic group.

Combined sample (n=911). The primary objective in performing the analyses described

above on the independent sample was to show that it was consistent with our prior work and therefore appropriate to collapse the independent and prior samples together, yielding a combined sample sufficiently large to analyze subsets parsed by sex. Having confirmed this, two-factor ANOVAs with diagnostic group (four levels) and sex (two levels) as fixed factors were performed for MSEL, VABS and ADOS scores to test for an effect of group, sex and a sex by diagnosis interaction. In addition, sex differences were tested in each diagnostic group independently with t-tests and correction for multiple comparisons.

Sex differences in reduced social attention by diagnostic group.

Combined sample (n=911). First, percentage of total looking time spent attending to any part of the approximately 1 minute long video (either AOI) between the four diagnostic groups (as an indication of general attentional and engagement group differences) was assessed with a 1-way ANOVA with Bonferroni correction. Then, a two factor ANCOVA with diagnostic group (four levels) and sex (two levels) as fixed factors, %Geo as dependent variable, and age at eye tracking as a covariate was used to test for effects of group, sex or a sex by diagnosis interaction in a large sample of toddler GeoPref eye tracking data. Partial eta-squared effect sizes are presented. In addition, sex differences were tested in each diagnostic group independently with t-tests and correction for multiple comparisons.

Clinical classification by reduced social attention scores, stratified by sex.

To assess the efficacy of %Geo scores as an ASD classifier, area under the ROC curve, sensitivity, specificity, and positive and negative predictive values were calculated for the combined sample (n=911), and for males and females separately. Spending 69% or more of looking time viewing the geometric AOI was the threshold for a positive test result (consistent with prior use of the GeoPref test) for females and males. Females are also presented with an alternate threshold that may optimize the test performance for detecting females.

As the ASD features group may reflect individuals just slightly below the categorical

diagnostic threshold, classification statistics are compared when the ASD features group scoring above the 69% positive test threshold for viewing geometric images are considered as either true positives or false positives, as in our past publication. However, the group ASD features do not meet DSM criteria for ASD, and including them as true positives did not improve classification performance, so these results are presented in the Supplemental Materials (Tables S2).

Characteristics of ASD subgroups defined by social attention, stratified by sex.

After identifying the female toddlers with the strongest preferences for geometric and social images using the 69% geometric cutoff, within ASD analyses compared the clinical characteristics (MSEL, VABS and ADOS) for these "GeoPref" and "SocPref" females. Comparisons of clinical scores (i.e. those of ADOS, MSEL and VABS) for effects of ASD subtype (GeoPref or SocPref), sex or a subtype X sex interaction were calculated with 2X2 factorial ANOVAs, after confirming that Levene's test of homogeneity of variance was nonsignificant. Partial eta-squared effect sizes are reported for significant effects. Post-hoc t-tests with Benjamini-Hochberg corrections for multiple comparisons and Cohen's d effect sizes were calculated for pairwise comparisons within each sex, reported one-tailed due the a priori hypothesis regarding the direction of effects that the GeoPref subtype would be more severely impaired.

Clinical phenotype differences based on familial risk and sex.

<u>High and low risk non-ASD sample (n=148).</u> High and low risk samples of non-ASD toddlers based on presence or absence of familial risk and an older sibling with ASD were identified from within the combined sample of n=911, in order to test the hypothesis that high risk females differ from other groups. Comparisons of clinical scores (i.e. those of ADOS, MSEL and VABS) for effects of risk, sex and a risk X sex interaction were calculated with 2X2

factorial ANOVAs, if Levene's test of homogeneity of variance was nonsignificant. For restricted repetitive behavior, significant heterogeneity of variance was found with Levene's test, therefore a nonparametric Kruskal-Wallis test was performed on the four groups (i.e. high risk males, high risk females, low risk males, low risk females). Partial eta-squared effect sizes are reported for significant effects. Post-hoc t-tests are reported one-tailed due to a priori hypotheses regarding the direction of effects, with Benjamini-Hochberg corrections for multiple comparisons.

Reduced social attention based on familial risk and sex

High and low risk non-ASD sample (n=148). To compare percentage of total looking time spent viewing the geometric AOI between groups (i.e. high risk males, high risk females, low risk males, low risk females) a 2X2 univariate ANCOVA was performed with the sex and risk status as factors, %Geo as the dependent variable, and age at eye tracking as a covariate. The relationships between social attention %Geo scores and ADOS, VABS and MSEL scores were assessed with Spearman's rank-order correlations as well.

Prior to selecting these analysis strategies, homogeneity of variance was confirmed with Levene's test. Also prior to these analyses, a 1-way ANOVA was used to check for differences between groups in total time attending to either area of interest during eye tracking, and there were no significant differences. A 1-way ANOVA was used to check for differences in age at eye tracking for the four groups, and again there were no significant differences. In alternative analyses, %Geo scores, which are not normally distributed, were first log transformed. This made no difference in the pattern of statistically significant and nonsignificant results, therefore, to maintain comparability to prior publications untransformed data are presented below.

Results

Replicating Clinical Phenotype Differences Based on Diagnostic Group

Independent sample (n=469). As expected, in the independent sample (n=469), a significant main effect of diagnosis was found for each comparison. The ASD group differed significantly from all other groups in each aspect of clinical phenotype assessed from the MSEL, ADOS and VABS, with means similar to our prior research (Moore et al., 2018). See Table 1.

Table 2.1: Clinical Phenotype, Independent Sample

(n=469)	1) ASD[a]	2) ASD Features	3) DD	4) TD	ASD vs 2), p=	ASD vs 3), p=	ASD vs 4), p=
N, Sex: M/F	94/28	24/3	139/39	81/61	n/a	n/a	n/a
Age at eye tracking, months: Mean (SD) [Range]	28.23 (8.5) [12-48]	25.11 (9.5) [12-44]	24.51 (9.8) [12-47]	25.03 (10.3) [12-48]	.773	.007	.044
ADOS[b] SA[c]	13.72 (4.6)	8.52 (5.3)	4.31 (3.3)	2.67 (2.1)	<.001	<.001	<.001
ADOS RRB[d]	4.17 (2.8)	2.56 (2.3)	.88 (1.2)	.32 (.63)	<.001	<.001	<.001
ADOS Total Score	17.88 (6.1)	11.07 (6.7)	5.19 (3.9)	2.99 (2.2)	<.001	<.001	<.001
MSEL[e] NV- DQ[f]	.80 (.20)	.97 (.20)	.96 (.27)	1.10 (.20)	.004	<.001	<.001
MSEL V- DQ[g]	.57 (.23)	.79 (.26)	.78 (.23)	1.04 (.13)	<.001	<.001	<.001
VABS ABC[h]	79.09 (9.7)	88.26 (8.8)	89.93 (10.1)	100.46 (8.5)	<.001	<.001	<.001

[a] See Methods for descriptions of diagnostic groups: ASD, ASD Features, Developmentally Delayed, and Typically Developing; [b] Autism Diagnostic Observation Schedule; [c] Social Affect score; [d] Restricted and Repetitive Behaviors score; [e] Mullen Scales of Early Learning; [f] Nonverbal Developmental Quotient; [g] Verbal Developmental Quotient; [h] Vineland Adaptive Behavior Scales, Adaptive Behavior Composite

Replicating Reduced Social Attention Differences Based on Diagnostic Group

Independent sample (n=469). Typically developing children spent more time attending to any part of the test video during eye tracking (52.7 seconds) than DD (48.6 seconds) and ASD (48.2 seconds) toddlers (p<.001 for both comparisons) and no other comparisons were

significant. Replicating previous findings, in the independent sample (n=469), the ASD group spent significantly more time viewing geometric images than any other group (for diagnostic group, $F_{(3,464)}$ =30.99, p<.001, partial eta-squared=.167). Toddlers with ASD spent 47.4% of looking time during eye tracking viewing geometric images; the group ASD features spent 29.4% of time; DD toddlers 25.3%; and TD toddlers 24.2% (ASD vs ASD features, p=.003; ASD vs DD, p<.001; ASD vs TD, p<.001).

Based on the consistency found between our prior publications of clinical phenotype data and GeoPref eye tracking data and what is reported above, the independent sample (n=469) and the previously published sample (n=442) are combined for the remainder of the analyses. This yields an overall sample size (n=911) large enough to make comparisons after splitting into subgroups such as ASD female and male toddlers, and high and low risk non-ASD female and male toddlers.

Sex Differences in Clinical Phenotype by Diagnostic Group

Combined sample (n=911). Within the overall sample, a few significant main effects of sex were found in clinical phenotype measures, for the ADOS and MSEL V-DQ (ADOS-SA, $F_{(1,896)}$ =5.35, p=.021, partial eta-squared=.006; ADOS total, $F_{(1,896)}$ =5.42, p=.020, partial eta-squared=.006; V-DQ, $F_{(1,859)}$ =9.33, p=.002, partial eta-squared=.011). Follow up t-test revealed no pairwise sex differences were significant for the groups ASD, ASD features, or TD, though ADOS and V-DQ scores were trending in non-ASD groups. A few sex differences in the DD group were found, but are probably not meaningful as they would not survive correction for multiple comparisons. This information is detailed in the supplement (Table S1). All main effects of diagnosis were strongly significant, very similar to the information presented in Table 1 for the independent sample, and in prior publications. There were no significant sex X diagnosis interactions in any clinical phenotype measures.

Sex Differences in Reduced Social Attention by Diagnostic Group

Combined sample (n=911). Typically developing children spent more time attending to any part of the test video during eye tracking (52.5 seconds) than DD (49.0 seconds) and ASD (48.8 seconds) toddlers ($p<.001$ for both comparisons), and no other comparisons were significant. While age at eye tracking ($F_{(1,902)}=76.51$, $p<.001$, partial eta-squared=.078) and diagnosis ($F_{(3,902)}=39.22$, $p<.001$, partial eta-squared=.115) contributed significantly to the ANCOVA model, there was no effect of sex ($F_{(1,902)}=.46$, $p=.498$, partial eta-squared=.001) or sex X diagnosis interaction ($F_{(3,902)}=.38$, $p=.767$, partial eta-squared=.001) in the large sample of n=911 toddler %Geo scores. For ASD males, %Geo scores were 45.32%; for ASD females, they were 46.74%. ASD features males scored 34.25% Geo, females scored 27.23%. DD males spent 26.48% of their time viewing the Geo AOI; DD females spent 24.15%. TD males' %Geo scores were 23.35%; females' were 23.96% (see Figure 2). None of these pairwise sex differences were significant.

Clinical Classification by Reduced Social Attention Scores, Stratified by Sex

ROC AUC, sensitivity and specificity from the GeoPref test were all highly similar between males and females, with AUC of .73 overall, .72 for males, and .75 for females (see Table 2). Positive predictive value was higher in males, but improved in females to 80% when using a cutoff specific to females (74% time viewing geometric images). Positive and negative predictive values are impacted by the rate of positive cases in the population. Therefore values presented reflect the test's performance on a pre-screened, at-risk sample, not on the general population, because as previously described our ascertainment method involves referral to the study based on low scores on a pediatrician administered developmental questionnaire (the CSBS). Including the ASD features group as true positives did not improve classification statistics (see Supplement).

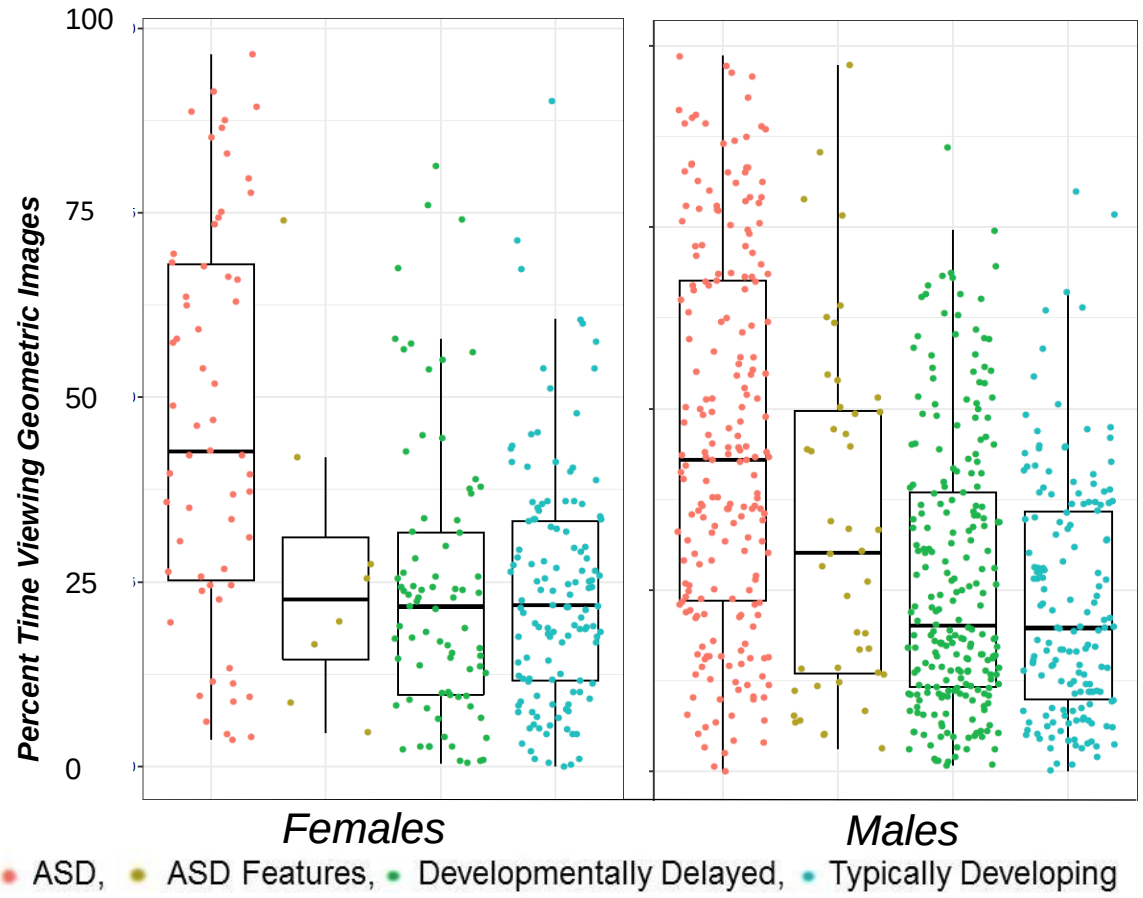

Figure 2.2: GeoPref Test Scores for Male and Female Toddlers (n=911). Scatterplots of individual subject % Geo scores for each diagnostic category and sex. Boxplots show median, range, and first and third quartiles. See Methods for descriptions of diagnostic groups.

Table 2.2: Clinical Classification Using Reduced Social Attention (%Geo) Scores

True positive = ASD only	All N=911 69% Geo threshold for test positive	Males N=638 69% Geo threshold for test positive	Females N=273 69% Geo threshold for test positive	Females N=273 74% Geo threshold for test positive
True positive	60	46	14	12
False negative	206	161	45	47
False positive	15	9	6	3
True negative	630	422	208	211
Sensitivity	23%	22%	24%	20%
Specificity	98%	98%	97%	99%
PPV[a]	80%	84%	70%	80%
NPV[b]	75%	72%	82%	82%
ROC AUC[c]	.73	.72	.75	.75

[a & b] positive and negative predictive values; [c] receiver operating characteristic area under the curve

Clinical Characteristics of ASD Subgroups Defined by Social Attention, Stratified by Sex

In male toddlers with ASD, 22% fall in the"GeoPref" subset with an average %Geo score of 82.23%, and 33% fall in the "SocPref" subset, with an average %Geo score of 16.83%. In females toddlers with ASD, 24% fall in the"GeoPref" subset with an average %Geo score of 82.66%, and 34% fall in the "SocPref" subset with an average %Geo score of 16.84%.

For the VABS, both MSEL scores, and ADOS social affect, 2X2 ANOVAs showed a significant main effect of GeoPref/SocPref subtype, plus no effect of sex, and no subtype by sex interaction (for VABS ABC, $F_{(1,142)} = 6.84$, p=.01, partial eta-squared=.046 for subtype; for MSEL NV-DQ, $F_{(1,138)} = 3.93$, p=.050, partial eta-squared=.028 for subtype; for MSEL V-DQ, $F_{(1,138)}= 10.77$, p=.001, partial eta-squared=.072 for subtype; and for ADOS SA, $F_{(1,142)}=11.76$, p=.001, partial eta-squared=.076 for subtype.) The ADOS total scores had a trending sex difference as well as a main effect of subtype and no interaction ($F_{(1,142)} =10.35$, p=.002, partial eta-squared= .068 for subtype, and $F_{(1,142)} =2.91$, p=.090, partial eta-squared=.020 for sex.) The ADOS RRB had no significant main effects or interaction.

Confirming prior findings, when comparing clinical scores from GeoPref and SocPref male toddlers with ASD, the GeoPref subtype is more severely impacted with regard to ADOS, V-DQ, and VABS ABC (NV-DQ is not significantly different), see Table 3. For female GeoPref and SocPref ASD subtype comparisons, V-DQ is significant with a large Cohen's d effect size (d=.91). No other pairwise comparisons of female subtypes are below p=.05 after correction for multiple comparisons. This is perhaps due to the small sample sizes for female GeoPref and SocPref subtypes, given that ADOS SA and ADOS total scores, MSEL NV-DQ and VABS all have Cohen's d effect sizes that are at least medium (d>.5) for females, and trending p values (from p=.07 to p=.10). There is no difference in ADOS RRB scores in females from the GeoPref and SocPref subsets.

Table 2.3: Clinical Phenotype Differences of ASD GeoPref and SocPref Subtypes by Sex

	ASD GeoPref[a] Subtype Male (n=46)	ASD SocPref Subtype Male (n=68)	ASD GeoPref Subtype Female (n=14)	ASD SocPref Subtype Female (n=20)	T tests (corrected p) and Cohen's d, Males	T tests (corrected p) and Cohen's d, Females
Age at eye tracking, months: Mean (SD) [Range]	29.54 (7.75) [18-48]	25.15 (7.85) [12-46]	29.64 (8.07) [20-48]	26.60 (7.89) [15-42]	p=.007, d=.56	p=.15, d=.38
ADOS[b] SA[c]	15.67 (4.08)	12.60 (4.65)	14.43 (5.58)	11.20 (4.88)	p=.002, d=.70	p=.07, d=.61
ADOS RRB[d]	4.62 (3.42)	3.55 (2.08)	3.50 (2.07)	3.30 (1.45)	p=.04, d=.38	p=.37, d=.11
ADOS Total	20.29 (5.93)	16.15 (5.82)	17.93 (6.51)	14.50 (5.73)	p=.003, d=.71	p=.08, d=.56
MSEL[e] NV- DQ[f]	.81 (.20)	.86 (.18)	.82 (.20)	.92 (.12)	p=.11, d=.25	p=.07, d=.61
MSEL V- DQ[g]	.53 (.25)	.64 (.20)	.53 (.18)	.73 (.24)	p=.02, d=.46	p=.03, d=.91
VABS[h] ABC	77.60 (9.9)	83.73 (10.5)	78.57 (7.44)	82.70 (8.81)	p=.005, d=.60	p=.10, d=.51

[a] See Methods for descriptions of GeoPref (geometric preference) and SocPref (social preference) ASD subtypes; [b] Autism Diagnostic Observation Schedule; [c] Social Affect score; [d] Restricted and Repetitive Behaviors score; [e] Mullen Scales of Early Learning; [f] Nonverbal Developmental Quotient; [g] Verbal Developmental Quotient; [h] Vineland Adaptive Behavior Scales, Adaptive Behavior Composite

Clinical Phenotype Differences Based on Familial Risk and Sex

High and low risk non-ASD sample (n=148). Analysis of sex X risk status for verbal developmental skills (MSEL V-DQ) revealed a significant effect for risk status ($F_{(1,144)}$ =4.73, p=.031, partial eta-squared=.032), a trend toward significance for sex ($F_{(1,144)}$ =3.52, p=.063, partial eta-squared=.024), and no sex X risk interaction. Post-hoc tests showed that high risk females had significantly higher verbal scores than low risk females ($t_{(68)}$ =2.09, p=.034) and a trending increase in verbal scores compared to high risk males ($t_{(55)}$ =1.70, p=.06). No differences were found between high and low risk males, or low risk males and low risk

females. Analysis of sex X risk status for non-verbal developmental skills (MSEL NV-DQ) showed no significant effects. Likewise, the analysis of sex X risk status for VABS Adaptive Behavior composite scores indicated no significant main effects.

Table 2.4: Clinical Phenotype Comparisons for non-ASD High/Low Risk Groups by Sex

	High Risk Females	**High Risk Males**	**Low Risk Females**	**Low Risk Males**
N, Sex: M/F	28	29	42	49
Age at eye tracking, months: Mean (SD) [Range]	22.29 (6.77) [12-38]	21.72 (9.01) [12-47]	21.12 (7.70) [12-40]	23.31 (9.24) [12-42]
ADOS[b] Social Affect	2.14 (2.01)	3.07 (2.49)	2.61 (2.16)	3.57 (2.69)
ADOS Restricted Repetitive Behavior	.25 (.44)	.79 (.98)	.19 (.67)	.49 (.77)
ADOS Total Score	2.39 (2.28)	3.86 (2.77)	2.81 (2.25)	4.06 (2.77)
MSEL[c] NV-DQ[d]	1.10 (.14)	1.08 (.17)	1.05 (.17)	1.07 (.30)
MSEL V-DQ[e]	1.05 (.17)	.97 (.18)	.96 (.17)	.92 (.21)
VABS ABC[f]	99.68 (10.38)	95.03 (10.04)	96.81 (12.65)	96.57 (11.50)

[a] See Methods for descriptions of high and low risk groups; [b] Autism Diagnostic Observation Schedule; [c] Mullen Scales of Early Learning; [d] Nonverbal Developmental Quotient; [e] Verbal Developmental Quotient; [f] Vineland Adaptive Behavior Scales, Adaptive Behavior Composite

ADOS total score analysis indicated an effect of sex (F (1,144) =9.98, p=.002, partial eta-squared=.065), and no significant effect of risk, or sex X risk interaction, with males scoring higher than females. Likewise, ADOS social affect scores showed a significant effect of sex (F (1,144) =5.40, p=.022, partial eta-squared=.036), and no effect of risk group or sex X risk interaction. Post-hoc tests showed that high risk females had significantly lower ADOS total scores than high risk males (t (53.7) =2.19, p=.041), and low risk females had a trend toward lower ADOS total scores than low risk males (t (89) =2.34, p=.055). However, high risk females did not have significantly lower ADOS total scores than low risk females. Post-hoc comparisons for ADOS social affect were trending with low risk females slightly lower than low risk males (t(89) =1.84, p=.058); other comparisons were not significant. ADOS RRB scores differed significantly as well

(Kruskal-Wallis H (3) = 14.56, p=.002). Post-hoc t-tests with equal variance not assumed showed high risk females had lower RRB scores than high risk males ($t_{(39.2)}$=2.72, p=.025), and low risk females trended toward lower RRB scores than low risk males ($t_{(89.0)}$ =1.99, p=.063).

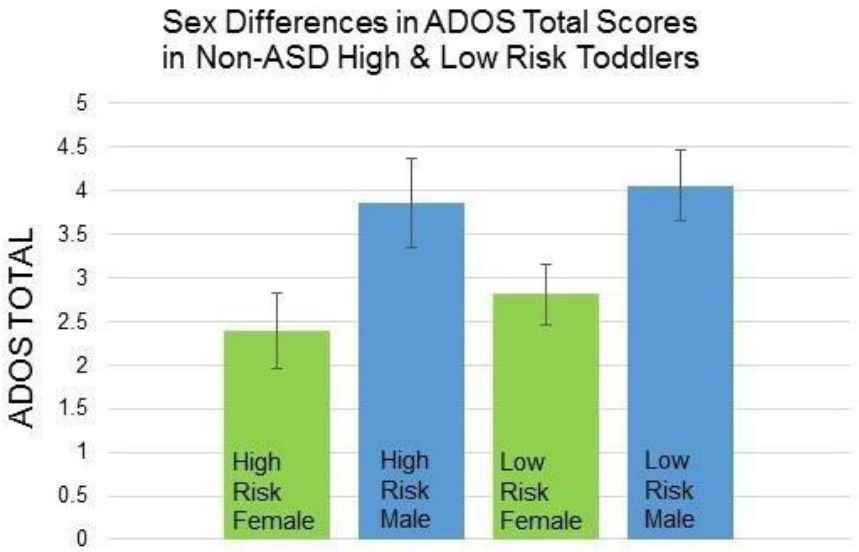

Figure 2.3: Autism Symptoms of Non-ASD Toddlers with and without ASD Older Siblings. Autism symptoms measured as sum of social affect and restricted, repetitive behavior ADOS scores. The high risk group have an older sibling with ASD; the low risk have no family history of ASD. Error bars are standard error of the mean.

Social Attention Based on Familial Risk and Sex

High and low risk non-ASD sample (n=148). Analysis of sex X risk status with age at eye tracking as a covariate showed no effect of sex, a trending effect of risk group ($F_{(1,143)}$=3.78, p=.054, partial eta-squared=.026), and a significant interaction between sex and risk group ($F_{(1,143)}$=6.32, p=.013, partial eta-squared=.042), where high risk females show the greatest social attention (lowest %Geo scores) and low risk females show the least social attention (See Figure 4). In an alternate analysis, V-DQ and NV-DQ were included in the model as covariates to control for differences in cognitive ability, but this made no difference in the pattern of statistically significant and nonsignificant results, suggesting the significant interaction is not driven by differences in cognitive developmental level.

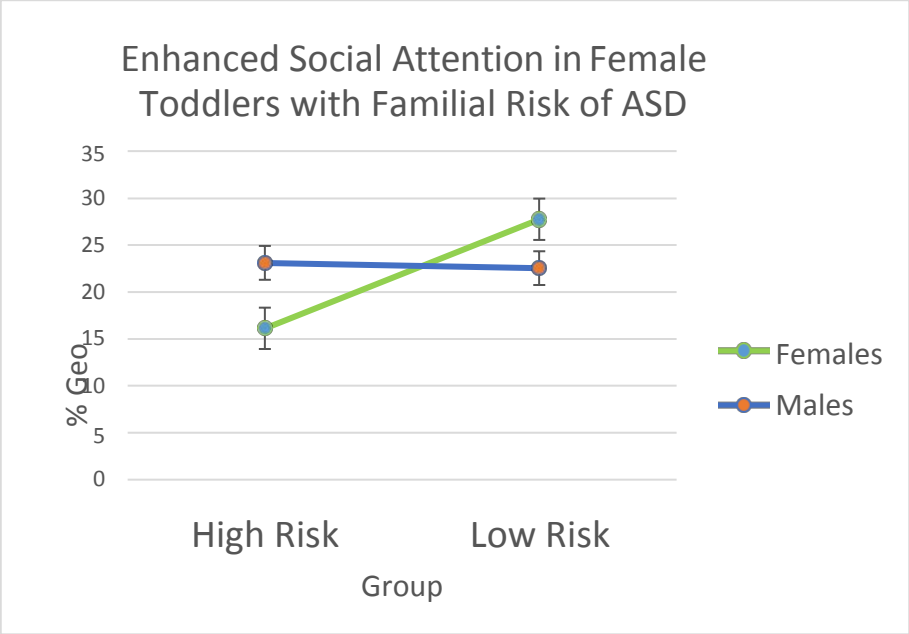

Figure 2.4: GeoPref Test Sex Differences in Non-ASD High and Low Risk Toddlers. Figure 4: % Geo scores from the GeoPref test (percent time viewing geometric videos out of total time viewing social or geometric videos) reflect reduced social attention. The high risk group have an older sibling with ASD; the low risk have no family history of ASD. Error bars are standard error of the mean.

Relationships Between Reduced Social Attention and Clinical Characteristics

For each of the four groups of interest, clinical scores presented above were related to %Geo scores with Spearman correlations, and a pattern of significant effects emerged. For ADOS total score and VABS adaptive behavior scores, reduced social attention was significantly correlated with phenotype for only one group, high risk males (%Geo to ADOS, rho=.519, p=.004; %Geo to VABS composite, rho= -.389, p=.037). Similarly, while no significant negative correlations were seen between %Geo scores and verbal and nonverbal developmental quotients, statistical trends were seen for only the high risk male group (%Geo to NV-DQ, rho= -.33, p=.08; %Geo to V-DQ, rho= -.347, p=.065). As our prior papers found %Geo scores to be correlated with ADOS scores for ASD and no other group, this finding is consistent with high risk males as ASD-like, or exhibiting a broader autism phenotype profile with regard to reduced social attention during the GeoPref test and ADOS scores. However,

this cannot be said for females.

Discussion

One aim of this study was to look for sex differences related to the GeoPref eye tracking test. There was, as expected, a strong effect of diagnosis on %Geo scores, indicating reduced social attention in ASD compared to all other diagnostic groups; however, there was no significant effect of sex, or sex by diagnosis interaction, suggesting reduced ASD social attention is indexed by the GeoPref test in both males and females. In addition, the performance of the GeoPref test as a classifier for examining toddlers with ASD based on eye tracking was very similar for each sex, though slightly better in females (area under ROC curve =.75) than in males (ROC AUC=.72). A female specific cutoff value (74% time viewing geometric stimuli) could be used to optimize the test's positive predictive value for females. Using that cutoff, this study confirms that the GeoPref test is similarly effective as a tool to identify both male and female toddlers with ASD, with moderate positive predictive value for both sexes (84% and 80%), and high specificity (98% and 99%) and low sensitivity (22% and 20%) for both sexes.

Similarly, when examining the clinical phenotypes (ADOS, MSEL and VABS) of ASD subtypes based on strong preference for geometric or social images during eye tracking, there was a main effect of group, and no significant effect of sex, or sex by subtype interaction, suggesting greater severity in the GeoPref subtype in both males and females, though this was clearer in males. The low sensitivity of the GeoPref test has been attributed to the test detecting a particular subset of ASD toddlers (the GeoPref subtype) who have elevated symptom severity shown in ADOS, MSEL and VABS scores. Here we show that in females with ASD, the GeoPref and SocPref subtypes are particularly characterized by a large difference in verbal ability (MSEL V-DQ), with females who attend primarily to the social images also significantly more advanced in language. The small sample sizes (n=14 and n=20) for the

two ASD subgroups in females is a limitation of this study.

Additionally, consistent with a prior study from a similar, screening referred sample (Reinhardt et al., 2015), there were no sex differences on any measure examined from ADOS, MSEL or VABS within the ASD group as a whole, or within the TD group. These data suggest a very similar presentation for ASD in male and female toddlers, with regard to symptom severity in the social communication and restricted, repetitive interests domains, and for verbal and nonverbal cognitive abilities and daily functioning. A limitation of this study, however, is that the assessments selected may be insensitive for detecting sex differences in restricted and repetitive behaviors, as only the ADOS specifically captures this domain and its range of scores is limited.

Another objective of the study was to test the hypothesis of enhanced social attention in high risk, non-ASD female toddlers. Our data confirm increased social attention during eye tracking in high risk females (16.1% time spent viewing geometric images) compared to low risk females (27.5% time spent viewing geometric images). This social attention interaction between sex and risk group in toddlers is similar to the prior report by Chawarska et al., of enhanced social attention in high risk females according to infant eye tracking during viewing of a naturalistic social scene. To further verify this finding, high risk female %Geo scores were compared to all other non-ASD female %Geo scores (without requiring an older sibling in the low risk female group: n=186, %Geo=25.4%), and this difference was also significant (p=.001).

High risk females also had the highest verbal developmental scores (significant even compared to low risk females), and lowest ADOS total scores of non-ASD groups in the risk by sex comparison. To further explore possible sex differences in non-ASD social attention, the proportion of toddlers with strong social preference (at least 69% time viewing social images of the GeoPref test) was compared across the entire group of N=645 non-ASD toddlers, and females were found to have a significantly higher rate of strong social preference

(72% of non-ASD females strongly preferred social images) compared to males (65% of non-ASD males strongly preferred social images, Chi-squared=3.84, p=.05). So, while the GeoPref test is similarly effective for detecting female and male ASD toddlers and no diagnostic groups had significant sex differences in %Geo scores, the GeoPref test does indicate subtle sex differences in social attention in non-ASD groups of toddlers.

Increased time spent viewing social scenes may indicate preference for the social, but could also reflect increased social vigilance or social anxiety in high risk compared to low risk females (Bishop, Veenstra-VanderWeele, & Sanders, 2016). This explanation was offered by Chawarska et al., and is consistent with reports that incomplete penetrance of ASD risk genes may present as anxiety in the female broader autism phenotype. For example, suggesting females can be biologically protected from complete autism penetrance, microdeletions of the gene SHANK1 lead to ASD in males, while the same mutation leads to anxiety but not ASD in females (Jeste & Geschwind, 2014; Sato et al., 2012). Consistent with this idea, during the GeoPref test, perhaps low risk females are more casually engaged, and their occasional looks to the geometric images reflect relaxed mind-wandering. High risk females may stay highly focused on the social scene due to being less relaxed and more anxiously vigilant. A follow-up study employing other measures (e.g. biometrics like galvanic skin response or heart rate) combined with eye tracking could investigate this.

These sex differences may reflect a “female protective effect”, or set of factors that protect females from developing ASD, and lead to the greater load of genetic risk factors found in females compared to males with ASD (Jacquemont et al., 2014). Whether and how the genetic female protective effect is related to behavioral traits found in females more often than males that predispose female toddlers toward engaging more with social information, like the increased processing of social information shown here, is so far unknown. Regardless of the cause of sex differences related to autism spectrum disorder, eye tracking studies with large

sample sizes such as this one are vital for making comparisons of subgroups that may be more homogenous than ASD as a whole, and for giving females with ASD full consideration in clinical research.

Acknowledgements

I would like to thank the UCSD Autism Center for making this work possible, specifically committee members Karen Pierce and Eric Courchesne for overarching support and guidance, Elizabeth Bacon for feedback on several drafts and frequent consulting on clinical and scientific matters, Mee-Kyoung Kwon and Srinivasa Nalabolu for eye tracking data management, students Orla Putnam and Sara Lee for assistance with data collection and pre-processing, psychologists Cindy Carter Barnes and Debra Cha for diagnostic assessments, Linda Lopez and Lauren Kupis for recruitment, scheduling and research assistant supervision, and Vahid Gazestani for help with R.

Supplemental Materials

Table 2.S1: Sex Differences in Clinical Characteristics

	Males	Females	T test, uncorrected p
ADOS Social Affect	***N=633***	***N=271***	
ASD	13.80 (4.6)	12.93 (4.6)	**.200**
ASD Features	8.77 (5.0)	6.88 (5.6)	**.335**
DD	4.31 (3.2)	3.76 (3.5)	**.197**
TD	2.48 (2.2)	2.13 (1.9)	**.145**
ADOS Restricted Repetitive	***N=633***	***N=271***	
ASD	4.00 (2.5)	3.46 (1.8)	**.122**
ASD Features	2.37 (1.9)	2.50 (2.6)	**.871**
DD	0.90 (1.2)	0.60 (1.0)	**.046**
TD	0.33 (0.7)	0.21 (0.5)	**.093**
ADOS Total Score	***N=633***	***N=271***	
ASD	17.80 (5.9)	16.39 (5.6)	**.105**
ASD Features	11.14 (6.2)	9.38 (6.4)	**.462**
DD	5.21 (3.7)	4.36 (4.0)	**.084**
TD	2.81 (2.4)	2.34 (2.1)	**.080**
MSEL Non-verbal Development	***N=612***	***N=255***	
ASD	.83 (.2)	.85 (.2)	**.458**
ASD Features	1.01 (.2)	1.07 (.2)	**.386**
DD	.97 (.3)	1.00 (.2)	**.491**
TD	1.12 (.2)	1.13 (.2)	**.664**
MSEL Verbal Development	***N=612***	***N=255***	
ASD	.61 (.2)	.64 (.2)	**.484**
ASD Features	.83 (.2)	.99 (.3)	**.083**
DD	.78 (.2)	.86 (.2)	**.002**
TD	1.05 (.1)	1.06 (.2)	**.298**
VABS ABC	***N=633***	***N=271***	
ASD	80.61 (10.1)	80.31 (10.0)	**.838**
ASD Features	89.74 (10.1)	94.63 (12.3)	**.232**
DD	90.37 (10.2)	91.82 (10.2)	**.281**
TD	100.73 (8.8)	101.47 (9.3)	**.494**

Table S1 above shows that the only pairwise sex differences that would survive robust correction for multiple comparisons in MSEL, ADOS, or VABS (though several others are trending) are Verbal-DQ scores in the Developmental Delay group. Note: Scores were excluded for subjects whose could not complete all testing in one visit and whose clinical testing was more than six months away from the date of eye tracking (n=7). Further, a subset of children with the greatest ability were administered the WPPSI rather than the MSEL, and their scores are also omitted (n=44). (ADOS - Autism Diagnostic Observation Schedule; MSEL - Mullen Scales of Early Learning; VABS - Vineland Adaptive Behavior Scales)

Table 2.S2: Inclusion of ASD features with %Geo >69% as true positive

69% Geo threshold for a positive test	N=911 True positive = ASD only	N=911 True positive = ASD and ASD Features	Males N=638 True positive = ASD only	Males True positive = ASD and ASD Features	Females N=273 True positive = ASD only	Females True positive = ASD and ASD Features
True positive	60	65	46	50	14	15
False negative	206	253	161	201	45	52
False positive	15	10	9	5	6	5
True negative	630	583	422	382	208	201
Sensitivity	23%	20%	22%	20%	24%	22%
Specificity	98%	98%	98%	99%	97%	98%
PPV[a]	80%	87%	84%	90%	70%	75%
NPV[b]	75%	70%	72%	66%	82%	79%
ROC AUC[c]	.73	.71	.72	.71	.75	.72

[a & b] positive and negative predictive values; [c] receiver operating characteristic area under the curve

Table S2 above shows that including subjects from the broader autism phenotype group “ASD features” with high %Geo scores as true positives rather than false positives does not improve classification statistics for the GeoPref test.

BIBLIOGRAPHY

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and emotion recognition from point light displays: an investigation of gender differences. *PLoS One, 6*(6), e20989. doi:10.1371/journal.pone.0020989

Alexander, G. M., Wilcox, T., & Woods, R. (2009). Sex differences in infants' visual interest in toys. *Arch Sex Behav, 38*(3), 427-433. doi:10.1007/s10508-008-9430-1

Andersson, G. W., Gillberg, C., & Miniscalco, C. (2013). Pre-school children with suspected autism spectrum disorders: do girls and boys have the same profiles? *Res Dev Disabil, 34*(1), 413-422. doi:10.1016/j.ridd.2012.08.025

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., . . . Dowling, N. F. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014 (vol 67, pg 1, 2018). *Mmwr-Morbidity and Mortality Weekly Report, 67*(19), 564-564.

Baron-Cohen, S., Bowen, D. C., Holt, R. J., Allison, C., Auyeung, B., Lombardo, M. V., . . . Lai, M. C. (2015). The "Reading the Mind in the Eyes" Test: Complete Absence of Typical Sex Difference in ~400 Men and Women with Autism. *PLoS One, 10*(8), e0136521. doi:10.1371/journal.pone.0136521

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry, 42*(2), 241-251.

Begeer, S., Mandell, D., Wijnker-Holmes, B., Venderbosch, S., Rem, D., Stekelenburg, F., & Koot, H. M. (2013). Sex differences in the timing of identification among children and adults with autism spectrum disorders. *J Autism Dev Disord, 43*(5), 1151-1156. doi:10.1007/s10803-012-1656-z

Bishop, S. L., Veenstra-VanderWeele, J., & Sanders, S. J. (2016). Attention Finally Being Paid to Girls at Risk of Autism. *J Am Acad Child Adolesc Psychiatry, 55*(3), 159-160. doi:10.1016/j.jaac.2015.12.013

Chawarska, K., Macari, S., Powell, K., DiNicola, L., & Shic, F. (2016). Enhanced Social Attention in Female Infant Siblings at Risk for Autism. *J Am Acad Child Adolesc Psychiatry, 55*(3), 188-195 e181. doi:10.1016/j.jaac.2015.11.016

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Res Dev Disabil, 48*, 79-93. doi:10.1016/j.ridd.2015.10.011

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., & Ahluwalia, J. (2000). Sex differences in human neonatal social perception. *Infant Behavior & Development, 23*(1), 113-118. doi:Doi 10.1016/S0163-6383(00)00032-1

Constantino, J. N. (2011). The quantitative nature of autistic social impairment. *Pediatr Res, 69*(5 Pt 2), 55R-62R. doi:10.1203/PDR.0b013e318212ec6e

Constantino, J. N. (2016). Data from the Baby Siblings Research Consortium confirm and specify the nature of the female protective effect in autism: A commentary on Messinger et al. *Mol Autism, 7*, 32. doi:10.1186/s13229-016-0092-x

Frazier, T. W., Georgiades, S., Bishop, S. L., & Hardan, A. Y. (2014). Behavioral and cognitive characteristics of females and males with autism in the Simons Simplex Collection. *J Am Acad Child Adolesc Psychiatry, 53*(3), 329-340 e321-323. doi:10.1016/j.jaac.2013.12.004

Frazier, T. W., & Hardan, A. Y. (2017). Equivalence of symptom dimensions in females and males with autism. *Autism, 21*(6), 749-759. doi:10.1177/1362361316660066

Frazier, T. W., Klingemier, E. W., Beukemann, M., Speer, L., Markowitz, L., Parikh, S., . . . Strauss, M. S. (2016). Development of an Objective Autism Risk Index Using Remote Eye Tracking. *Journal of the American Academy of Child and Adolescent Psychiatry, 55*(4), 301-309. doi:10.1016/j.jaac.2016.01.011

Fujioka, T., Inohara, K., Okamoto, Y., Masuya, Y., Ishitobi, M., Saito, D. N., . . . Kosaka, H. (2016). Gazefinder as a clinical supplementary tool for discriminating between autism spectrum disorder and typical development in male adolescents and adults. *Mol Autism, 7*, 19. doi:10.1186/s13229-016-0083-y

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *J Autism Dev Disord, 37*(4), 613-627. doi:10.1007/s10803-006-0280-1

Halladay, A. K., Bishop, S., Constantino, J. N., Daniels, A. M., Koenig, K., Palmer, K., . . Szatmari, P. (2015). Sex and gender differences in autism spectrum disorder: summarizing evidence gaps & identifying emerging areas of priority. *Mol Autism, 6*, 36. doi:10.1186/s13229-015-0019- y

Harrop, C., Green, J., Hudry, K., & Consortium, P. (2017). Play complexity and toy engagement in preschoolers with autism spectrum disorder: Do girls and boys differ? *Autism, 21*(1), 37-50. doi:10.1177/1362361315622410

Harrop, C., Gulsrud, A., & Kasari, C. (2015). Does Gender Moderate Core Deficits in ASD? An Investigation into Restricted and Repetitive Behaviors in Girls and Boys with ASD. *J Autism Dev Disord, 45*(11), 3644-3655. doi:10.1007/s10803-015-2511-9

Harrop, C., Jones, D., Zheng, S., Nowell, S., Boyd, B. A., & Sasson, N. (2018). Circumscribed Interests and Attention in Autism: The Role of Biological Sex. *J Autism Dev Disord*. doi:10.1007/s10803-018-3612-z

Hartley, S. L., & Sikora, D. M. (2009). Sex differences in autism spectrum disorder: an examination of developmental functioning, autistic symptoms, and coexisting behavior problems in toddlers. *J Autism Dev Disord, 39*(12), 1715-1722. doi:10.1007/s10803-009-0810-8

Holt, R. J., Chura, L. R., Lai, M. C., Suckling, J., von dem Hagen, E., Calder, A. J., . . . Spencer, M. D. (2014). 'Reading the Mind in the Eyes': an fMRI study of adolescents with autism and their siblings. *Psychol Med, 44*(15), 3215-3227. doi:10.1017/S0033291714000233

Hull, L., Mandy, W., & Petrides, K. V. (2017). Behavioural and cognitive sex/gender differences in

autism spectrum condition and typically developing males and females. *Autism, 21*(6), 706-727. doi:10.1177/1362361316669087

Jacquemont, S., Coe, B. P., Hersch, M., Duyzend, M. H., Krumm, N., Bergmann, S., . . . Eichler, E. E. (2014). A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. *Am J Hum Genet, 94*(3), 415-425. doi:10.1016/j.ajhg.2014.02.001

Jeste, S. S., & Geschwind, D. H. (2014). Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol, 10*(2), 74-81. doi:10.1038/nrneurol.2013.278

Kreiser, N. L., & White, S. W. (2014). ASD in Females: Are We Overstating the Gender Difference in Diagnosis? *Clinical Child and Family Psychology Review, 17*(1), 67-84. doi:10.1007/s10567-013-0148-9

Lai, M. C., Baron-Cohen, S., & Buxbaum, J. D. (2015). Understanding autism in the light of sex/gender. *Mol Autism, 6*, 24. doi:10.1186/s13229-015-0021-4

Lai, M. C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/gender differences and autism: setting the scene for future research. *J Am Acad Child Adolesc Psychiatry, 54*(1), 11-24. doi:10.1016/j.jaac.2014.10.003

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child and Adolescent Psychiatry, 56*(6), 466-474. doi:10.1016/j.jaac.2017.03.013

Lord, C., & Schopler, E. (1985). Differences in sex ratios in autism as a function of measured intelligence. *J Autism Dev Disord, 15*(2), 185-193.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., . . . Lord, C. (2009). The Autism Diagnostic Observation Schedule-toddler module: a new module of a standardized diagnostic measure for autism spectrum disorders. *J Autism Dev Disord, 39*(9), 1305-1320. doi:10.1007/s10803-009-0746-z

MacDonald, R., Parry-Cruwys, D., Dupere, S., & Ahearn, W. (2014). Assessing progress and outcome of early intensive behavioral intervention for toddlers with autism. *Res Dev Disabil, 35*(12), 3632-3644. doi:10.1016/j.ridd.2014.08.036

Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2012). Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. *J Autism Dev Disord, 42*(7), 1304-1313. doi:10.1007/s10803-011-1356-0

Messinger, D. S., Young, G. S., Webb, S. J., Ozonoff, S., Bryson, S. E., Carter, A., . . . Zwaigenbaum, L. (2015). Early sex differences are not autism-specific: A Baby Siblings Research Consortium (BSRC) study. *Mol Autism, 6*, 32. doi:10.1186/s13229-015-0027-y

Messinger, D. S., Young, G. S., Webb, S. J., Ozonoff, S., Bryson, S. E., Carter, A., . . . Zwaigenbaum, L. (2016). Commentary: sex difference differences? A reply to Constantino. *Mol Autism, 7*, 31. doi:10.1186/s13229-016-0093-9

Moore, A., Wozniak, M., Yousef, A., Barnes, C. C., Cha, D., Courchesne, E., & Pierce, K. (2018). The geometric preference subtype in ASD: identifying a consistent, early-emerging phenomenon through eye tracking. *Mol Autism, 9*, 19. doi:10.1186/s13229-018-0202-z

Murias, M., Major, S., Davlantis, K., Franz, L., Harris, A., Rardin, B., . . . Dawson, G. (2018). Validation of eye-tracking measures of social attention as a potential biomarker for autism clinical trials. *Autism Res, 11*(1), 166-174. doi:10.1002/aur.1894

Pierce, K., Carter, C., Weinfeld, M., Desmond, J., Hazin, R., Bjork, R., & Gallagher, N. (2011). Detecting, studying, and treating autism early: the one-year well-baby check-up approach. *J Pediatr, 159*(3), 458-465 e451-456. doi:10.1016/j.jpeds.2011.02.036

Pierce, K., Conant, D., Hazin, R., Stoner, R., & Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. *Arch Gen Psychiatry, 68*(1), 101-109. doi:10.1001/archgenpsychiatry.2010.113

Pierce, K., Marinero, S., Hazin, R., McKenna, B., Barnes, C. C., & Malige, A. (2016). Eye Tracking Reveals Abnormal Visual Preference for Geometric Images as an Early Biomarker of an Autism Spectrum Disorder Subtype Associated With Increased Symptom Severity. *Biol Psychiatry, 79*(8), 657-666. doi:10.1016/j.biopsych.2015.03.032

Ratto, A. B., Kenworthy, L., Yerys, B. E., Bascom, J., Wieckowski, A. T., White, S. W., . . . Anthony, L. G. (2018). What About the Girls? Sex-Based Differences in Autistic Traits and Adaptive Skills. *J Autism Dev Disord, 48*(5), 1698-1711. doi:10.1007/s10803-017-3413-9

Reinhardt, V. P., Wetherby, A. M., Schatschneider, C., & Lord, C. (2015). Examination of sex differences in a large sample of young children with autism spectrum disorder and typical development. *J Autism Dev Disord, 45*(3), 697-706. doi:10.1007/s10803-014-2223-6

Sato, D., Lionel, A. C., Leblond, C. S., Prasad, A., Pinto, D., Walker, S., . . . Scherer, S. W. (2012). SHANK1 Deletion in Males with Autism Spectrum Disorder. *Am J Hum Genet, 90*(5), 879-887. doi: 10. 1016/j.ajhg.2012.03.017

Shattuck, P. T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D. S., Wiggins, L., . . . Cuniff, C. (2009). Timing of identification among children with an autism spectrum disorder: findings from a population-based surveillance study. *J Am Acad Child Adolesc Psychiatry, 48*(5), 474-483. doi:10.1097/CHI.0b013e31819b3848

Szatmari, P., Liu, X. Q., Goldberg, J., Zwaigenbaum, L., Paterson, A. D., Woodbury-Smith, M., . . Thompson, A. (2012). Sex differences in repetitive stereotyped behaviors in autism: implications for genetic liability. *Am J Med Genet B Neuropsychiatr Genet, 159B*(1), 5-12. doi: 10.1002/ ajmg.b.31238

Tsai, L. Y., & Beisler, J. M. (1983). The development of sex differences in infantile autism. *Br J Psychiatry, 142*, 373-378.

Van Wijngaarden-Cremers, P. J., van Eeten, E., Groen, W. B., Van Deurzen, P. A., Oosterling, I. J., & Van der Gaag, R. J. (2014). Gender and age differences in the core triad of impairments in autism spectrum disorders: a systematic review and meta-analysis. *J Autism Dev Disord, 44*(3), 627-635. doi:10.1007/s10803-013-1913-9

Vargas-Cuentas, N. I., Roman-Gonzalez, A., Gilman, R. H., Barrientos, F., Ting, J., Hidalgo, D., . . . Zimic, M. (2017). Developing an eye-tracking algorithm as a potential tool for early diagnosis of autism spectrum disorder in children. *PLoS One, 12*(11), e0188826. doi:10.1371/journal.pone.0188826

Vivanti, G., Dissanayake, C., & Victorian, A. T. (2016). Outcome for Children Receiving the Early Start Denver Model Before and After 48 Months. *J Autism Dev Disord, 46*(7), 2441-2449. doi:10.1007/s10803-016-2777-6

Volkmar, F. R., Szatmari, P., & Sparrow, S. S. (1993). Sex differences in pervasive developmental disorders. *J Autism Dev Disord, 23*(4), 579-591.

Werling, D. M. (2016). The role of sex-differential biology in risk for autism spectrum disorder. *Biol Sex Differ, 7*, 58. doi:10.1186/s13293-016-0112-8

Werling, D. M., & Geschwind, D. H. (2013). Understanding sex bias in autism spectrum disorder. *Proc Natl Acad Sci U S A, 110*(13), 4868-4869. doi:10.1073/pnas.1301602110

Wetherby, A. M., Brosnan-Maddox, S., Peace, V., & Newton, L. (2008). Validation of the Infant-Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism, 12*(5), 487-511. doi:10.1177/1362361308094501

Zwaigenbaum, L., Bryson, S. E., Szatmari, P., Brian, J., Smith, I. M., Roberts, W., . . . Roncadin, C. (2012). Sex differences in children with autism spectrum disorder identified within a high-risk infant cohort. *J Autism Dev Disord, 42*(12), 2585-2596. doi:10.1007/s10803-012-1515-y

CHAPTER 3

Moore et al. Molecular Autism (2018) 9:19
https://doi.org/10.1186/s13229-018-0202-z

Molecular Autism

RESEARCH Open Access

The geometric preference subtype in ASD: identifying a consistent, early-emerging phenomenon through eye tracking

Adrienne Moore*, Madeline Wozniak, Andrew Yousef, Cindy Carter Barnes, Debra Cha, Eric Courchesne and Karen Pierce

Abstract

Background: The wide range of ability and disability in ASD creates a need for tools that parse the phenotypic heterogeneity into meaningful subtypes. Using eye tracking, our past studies revealed that when presented with social and geometric images, a subset of ASD toddlers preferred viewing geometric images, and these toddlers also had greater symptom severity than ASD toddlers with greater social attention. This study tests whether this "GeoPref test" effect would generalize across different social stimuli.

Methods: Two hundred and twenty-seven toddlers (76 ASD) watched a 90-s video, the Complex Social GeoPref test, of dynamic geometric images paired with social images of children interacting and moving. Proportion of visual fixation time and number of saccades per second to both images were calculated. To allow for cross-paradigm comparisons, a subset of 126 toddlers also participated in the original GeoPref test. Measures of cognitive and social functioning (MSEL, ADOS, VABS) were collected and related to eye tracking data. To examine utility as a diagnostic indicator to detect ASD toddlers, validation statistics (e.g., sensitivity, specificity, ROC, AUC) were calculated for the Complex Social GeoPref test alone and when combined with the original GeoPref test.

Results: ASD toddlers spent a significantly greater amount of time viewing geometric images than any other diagnostic group. Fixation patterns from ASD toddlers who participated in both tests revealed a significant correlation, supporting the idea that these tests identify a phenotypically meaningful ASD subgroup. Combined use of both original and Complex Social GeoPref tests identified a subgroup of about 1 in 3 ASD toddlers from the "GeoPref" subtype (sensitivity 35%, specificity 94%, AUC 0.75.) Replicating our previous studies, more time looking at geometric images was associated with significantly greater ADOS symptom severity.

Conclusions: Regardless of the complexity of the social images used (low in the original GeoPref test vs high in the new Complex Social GeoPref test), eye tracking of toddlers can accurately identify a specific ASD "GeoPref" subtype with elevated symptom severity. The GeoPref tests are predictive of ASD at the individual subject level and thus potentially useful for various clinical applications (e.g., early identification, prognosis, or development of subtype-specific treatments).

Keywords: Eye tracking, Autism spectrum disorder, Early identification, Social attention, Geometric preference

* Correspondence: armoore@ucsd.edu
Autism Center of Excellence, Department of Neurosciences, University of California San Diego, La Jolla, CA, USA

Background

Autism spectrum disorder (ASD) encompasses a heterogeneous collection of phenotypes. Some individuals with ASD are highly capable, verbally fluent individuals who view their autism as a benign difference requiring an increase in tolerance and acceptance from the neurotypical community rather than a cure [1]. Others with ASD are severely impaired with minimal ability for self-care or for communicating their perspectives or needs [2–4]. It may be possible to maximize impact of treatment for those with particularly challenging forms of ASD by intervening early in the development of symptoms [5–7]. Neurobiological differences between people who will go on to be diagnosed with ASD and those who will not have been traced back to even prenatal stages of development [8–10]. Differences in behavioral presentation at the group level between children who will and will not go on to be diagnosed with ASD have been found as early as 6 months [11–13]. However, according to recent Centers for Disease Control and Prevention reporting, most children on the autism spectrum in the USA are not diagnosed until after the age of 4 [14]. The development of effective tests that can reliably identify in their infancy which individuals will go on to be diagnosed with autism, and whether that autism will ultimately be mild or severe, is in its very early stages.

Clinician judgments of observed behavior, though vulnerable to subjective bias, remain the gold standard for ASD identification [15]. This state of affairs persists despite widespread acceptance that the origins of ASD are neurobiological and that therefore the development of highly objective tests should be achievable [16]. Eye tracking is a methodology with great potential clinical utility for screening, diagnosis, and early detection of ASD [17]. It is objective, quantitative, non-invasive, relatively inexpensive and easy to use, and appropriate for very young infants and many levels of functioning [18, 19]. Moreover, while basic oculomotor functioning has been shown not to differ in fundamental ways [19] between ASD and controls (although see [20, 21] for notable differences in spontaneous fixation durations and attentional disengagement), patterns of viewing socially relevant information reveal the phenotypic differences between ASD and typical development [22]. When paired with stimuli and tasks that have been well explored by the field of neuroscience, eye tracking may move us toward clinical approaches grounded in knowledge of the disrupted neural circuitry of ASD, with the goal of improved treatment impact [23].

Eye tracking studies of toddlers and young children with ASD have reported less time attending to biological motion [24], less attention to people's heads and more to bodies [25], less time viewing people and faces within a complex scene [13], and, when viewing faces, less time spent viewing the key feature components [26]. Young children with ASD also exhibit atypical gaze-following behavior during eye tracking paradigms [27] which is important because gaze-following is a key precursor to the development of joint attention [28]. Joint attention skills are critically associated with language acquisition in typically developing children [29] and with language and social deficits in ASD [30]. Though it has not been demonstrated, the social differences and difficulties of adults with ASD could possibly be influenced by the long-term, cumulative impact of this abnormal visual attention to what is socially meaningful during development [31–33]. Abnormal non-social attentional components (e.g., disengagement) likely add complexity to this explanation as well [34]. Intervention studies focused on improving joint attention skills have yielded promising results thus far, suggesting it may be feasible to alter this course of events as ASD unfolds across childhood [35].

Despite the many insights into ASD development stemming from eye tracking research, difficultly when comparing results from different eye tracking studies of ASD toddlers has been noted [22]. This is in part because seemingly minor changes to the stimuli presented may alter the results considerably. For example, in separate studies, Jones [31] and Chawarska [36] presented video stimuli to ASD toddlers of similar ages (mean age 2.1 years, standard deviation .65 years, and the 13- to 25-month age range, respectively). Both studies included a complex stimulus background with toys and other objects, in front of which was a centrally located female actress who looked directly into the camera while trying to attract attention with child-directed speech. However, Jones (2008) found increased looking to the mouth region and decreased looking to the eye region in ASD children compared to contrast groups. Chawarska, on the other hand, found decreased looking time to the face and specifically to the mouth region and increased looking to the hands in ASD children, yet no differences in eye region fixation per se, compared to contrast groups. There are various possible explanations for this discrepancy, and several authors have previously commented on it [22, 37]. Regardless of the cause, the sharp inconsistency of findings between such studies suggests that results when eye tracking toddlers with ASD can be very sensitive and may not generalize robustly or replicate unless many factors are controlled.

In contrast, the current eye tracking study examines whether the GeoPref test effect is robust against changes to the social images presented in a conceptual replication of the "GeoPref" subtype effect identified in our previous work (see Fig. 1). Specifically, in 2011, Pierce et al. reported eye tracking data from a preferential looking task showing that preference for viewing geometric rather than social stimuli is a risk factor for an autism diagnosis in toddlers [38]. In 2016, Pierce et al. reported that this Geometric Preference (GeoPref) test identifies

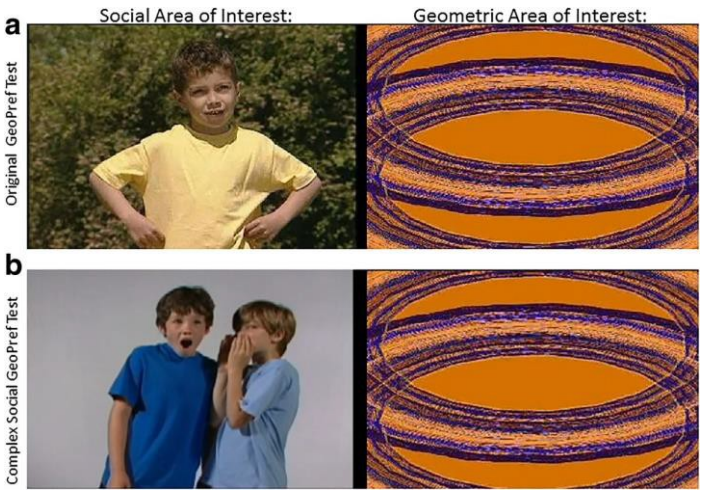

Fig. 1 Examples of stimuli. a Sample image from the original GeoPref test which consists of a 60-s video composed of 28 scenes, 26 of which involve one individual's movement (while the other two scenes include twins moving side by side). ©2003 Gaiam Americas, Inc., Courtesy of Gaiam Americas, Inc. As each social scene switches to a new actor, the paired geometric scene switches simultaneously to differen t colors and moving shapes. b Sample image from the Complex Social GeoPref test which consists of a 90-s video composed of nine scenes, five of which include two children interacting and four of which include one child moving enthusiastically. As each social scene changes, the pai red geometric scene also switches simultaneously to a different color and moving shape, as in the original GeoPref test

an ASD subtype with increased symptom severity compared to ASD children who preferentially view social images [39]. Individuals in the GeoPref ASD subtype, who spent more than 69% of looking time viewing geometric stimuli, had higher ADOS scores, lower MSEL receptive and expressive language and visual reception scores, and lower Vineland scores for adaptive behavior.

The current study tests the Geometric Preference phenomenon identified previously by varying the social stimulus presentation's total length, scene length, and complexity of social interactions. The original study shows full body or large, dance-like movements anduniformly positive affect. The current study depicts a broader range of expressed emotions including surprise and anger, as well as happiness, and shows socially meaningful but physically more subtle actions like whispering in another's ear, hugging, and one child sticking out her tongue at another. As in the original test, these social stimuli portray biological motion and faces (though with less biological motion and more varied facial expressions), keeping the paradigm closely linked to stimuli that have been often used by cognitive neuroscience in attempts to map the social brain [40–42]. These complex social vignettes unfold more gradually than the actions in the original stimuli; therefore, the stimulus video is longer overall and composed of longer individual scenes. The geometric stimuli were not altered, other than by selecting a subset and extending the duration of presentation per scene to match the durations to the social stimuli. This was done so that by isolating the social variables only, we could conclude that changes in the pattern of responses to viewing the stimuli were due to the social scene manipulation, thus avoiding any confounds. Because toddlers with ASD are more likely to have a reduced interest in social stimuli, we considered the fact that the sensitivity of the test (which was around 23% for the original GeoPref test) might improve if we altered the complexity of the social stimuli. That is, we predicted that a greater percentage of ASD toddlers may find the social side uninteresting (and would thus fixate on the geometric images instead) if it were made more complex. However, for typical toddlers, it may increase their interest in the social side if social interactions were depicted, which would potentially increase group differences.

In a meta-analysis of 38 articles comparing ASD and TD children using eye tracking [43], Chita-Tegmark reports that increasing the social content of stimuli by showing more than one person is the factor that best reveals the differences in social attention between ASD and TD groups. We tested the hypothesis that the original GeoPref test identifies a stable subtype of autism characterized by robust patterns of decreased social attention and increased attention to geometric repetition and therefore that the Complex Social GeoPref test should generally replicate the findings of the original GeoPref test, perhaps with amplified effects due to changes to the social stimuli used. That is, ASD children were predicted on average to have greater fixation times on geometric images than contrast groups, and above

some threshold, all children with sufficiently high fixation time on geometric images were predicted to be ASD children. Further, ASD children who complete both tests within the appropriate age range were predicted to have fairly stable scores. Additionally, the ASD children with greatest fixation times on geometric images, the GeoPref ASD subtype, were predicted to also have worse cognitive, language, and social skills based on MSEL, ADOS, and VABS scores compared with ASD children with the least geometric fixation times (the ASD "SocPref" subtype).

Methods

Participant recruitment

Two hundred and seventy toddlers enrolled in and completed this study. Of the 270, 43 were excluded from data analysis for reasons detailed in Additional file 1: Figure S1 (e.g., vision abnormality, tantrum during eye tracking), leaving a final study sample of 227 toddlers. Their ages ranged from 12 to 48 months (mean 29.5 months, standard deviation 9.5). Two hundred and eleven of the 227 subjects in the present study (93%) were new and non-overlapping with our past two eye tracking papers [38, 39]. Of the 227 participants, 126 completed both the Complex Social GeoPref eye tracking test newly described herein and the original GeoPref test described in previous publications [38, 39, 44]. Sixty-eight of these 126 (54%) completed the original GeoPref test first, and 58 (46%) completed the Complex Social GeoPref test first, and no age differences were found between groups at either time point. The remaining 101 subjects completed the new Complex Social GeoPref test but not the original GeoPref test.

All diagnostic, psychometric and eye tracking tests took place at the University of California San Diego Autism Center. During data collection time periods, any child receiving an autism evaluation, regardless of referral source, was included in eye tracking testing. Fifty-four percent of the sample of 227 were referred to us by their pediatrician who participates in our general population-based screening method called the 1-Year Well-Baby Check-Up Approach [44]. This allows for the prospective study of ASD, as well as global developmental and language delay or other delays, beginning as early as 12 months, typically based on a toddler's failure of the CSBS-DP Infant-Toddler Checklist [45]. Occasionally, a child is referred by a participating pediatrician between ages 2 and 3 so the CSBS is no longer applicable, or because there is concern regarding the child's development despite a passing score on the CSBS questionnaire. The remaining 46% of subjects were not referred by their pediatricians. These participants either self-referred due to parental concern about their child's development, or participated as controls. Though they were not referred after pediatrician screening for developmental delays, these children received identical testing to the screening referred group during their evaluations at the UCSD Autism Center. ASD children comprise 38% of the group referred through pediatrician screening and 29% of the self-referred group, and this difference falls short of statistical significance (chi-squared = 2.09, p = .15). Mean age at eye tracking for the pediatrician screening referred group was 29.9 months; mean age for the self-referred group was 29.0 months at eye tracking.

Diagnostic and psychometric assessments

At each visit, assessments were administered at UCSD Autism Center by PhD-level licensed clinical psychologists and included the Autism Diagnostic Observation Schedule (ADOS) module T, 1, or 2 [46], Mullen Scales of Early Learning (MSEL) [47], and Vineland Adaptive Behavior Scales (VABS) [48]; additional family and medical histories were also obtained. Toddlers who participated when younger than age 30 months were longitudinally tracked and diagnostically evaluated every 6–12 months until age 3 years when a final diagnosis was given. Any child receiving an evaluation during the data collection time period was administered eye tracking for this study, regardless of whether their visit was an intake appointment, a follow-up, or a final diagnostic appointment. Table 1 presents characteristics of the sample.

The study sample consisted of six discrete diagnostic groups of toddlers: 76 ASD, 11 ASD features, 56 DD, 51 TD, 22 Other, and 11 TypSib. The *ASD* group included toddlers who met DSM criteria for Autistic Disorder or PDD-NOS (DSM IV) or ASD (DSM V) at their final diagnostic evaluation. The *ASD features* group had significant ASD symptoms and/or elevated ADOS scores during at least one evaluation but did not meet full criteria for ASD at their final longitudinal evaluation. The *DD* group included transient and persistent language delay and global developmental delay determined by MSEL scores. The *TD* group included "type 1 errors," children who failed the CSBS screening at a pediatric visit but tested within typical levels on ADOS, MSEL, and VABS during their evaluations, as well as typically developing toddlers who both passed the CSBS and tested within the typical range on ADOS, MSEL, and VABS tests during their evaluations. In the *TypSib* group were unaffected toddlers with siblings with ASD who tested within the typical range during their evaluations. In the *Other* group were toddlers with a wide array of other conditions such as social anxiety or a tic disorder. For this study, 83% of the overall sample received a final diagnostic assessment at 30 months or older (mean age 38.3 months). The remaining 17% (*13 ASD, 18 DD, 4 TD, 2 TypSib, 1 Other*) were assigned to a diagnostic

Table 1 Participant characteristics of overall sample

	1) ASD[a]	2) ASD feat.	3) DD	4) TD	5) Other	6) Typical sibling ASD	ASD vs 2), *p*=	ASD vs 3), *p*=	ASD vs 4), *p*=	ASD vs 5), *p*=	ASD vs 6), *p*=
Sex, M/F	70/6	10/1	36/20	30/21	11/11	4/7	n/a	n/a	n/a	n/a	n/a
Age at eye tracking, months Mean (SD) [range]	30.0 (8.8) [12.1–47.4]	31.9 (8.9) [15.8–40.7]	26.8 (9.5) [12.4–46.0]	29.7 (9.5) [12.9–47.5]	33.6 (10.3) [13.1–47.7]	27.8 (11.2) [12.2–44.6]	NS	NS	NS	NS	NS
MSEL AE[b] scores/true age											
Visual reception	.79 (.18)	.94 (.24)	.93 (.19)	1.16 (.18)	1.05 (.20)	1.17 (.16)	NS	< .001	< .001	< .001	< .001
Fine motor	.79 (.18)	.90 (.13)	.96 (.18)	1.04 (.13)	.95 (.20)	1.04 (.16)	NS	< .001	< .001	< .01	< .001
Receptive language	.58 (.28)	.82 (.24)	.79 (.24)	1.10 (.16)	1.0 (.23)	1.03 (.15)	< .05	< .001	< .001	< .001	< .001
Expressive language	.60 (.26)	.72 (.20)	.64 (.23)	1.04 (.17)	.99 (.25)	1.04 (.13)	NS	NS	< .001	< .001	< .001
VABS[c] standard scores											
Communication	75.2 (18.8)	90.8 (17.2)	84.6 (15.4)	104.1 (9.8)	98.8 (12.0)	101.5 (5.1)	< .05	< .05	< .001	< .001	< .001
Daily living	82.4 (15.5)	99.2 (11.7)	92.2 (13.2)	101.1 (9.3)	96.7 (15.2)	98.4 (11.6)	< .005	= .001	< .001	< .001	< .005
Socialization	80.1 (16.1)	96.0 (11.3)	96.4 (11.2)	102.7 (9.2)	97.4 (13.7)	104.4 (6.6)	< .005	< .001	< .001	< .001	< .001
Motor skills	87.8 (17.3)	96.6 (10.4)	91.0 (13.4)	99.7 (9.6)	94.8 (13.3)	99.9 (6.5)	NS	NS	< .001	NS	NS
Composite score	78.8 (15.8)	94.6 (11.7)	88.6 (11.2)	102.7 (10.4)	96.1 (13.1)	100.9 (5.7)	< .005	< .001	< .001	< .001	< .001
ADOS[d] module T, 1, or 2											
SA[e]	13.2 (4.2)	6.2 (5.5)	4.3 (3.1)	2.5 (1.7)	4.3 (3.7)	3.1 (2.2)	< .001	< .001	< .001	< .001	< .001
RRB[f]	3.7 (2.0)	2.0 (1.5)	0.8 (1.8)	0.4 (0.8)	0.4 (0.6)	0.5 (0.5)	< .05	< .001	< .001	< .001	< .001
Total score	16.9 (5.2)	8.2 (6.0)	5.2 (3.8)	2.9 (2.0)	4.7 (4.0)	3.6 (2.0)	< .001	< .001	< .001	< .001	< .001

[a]See text for descriptions of diagnostic groups ASD, ASD Feat., DD, TD, Other, Typical Sibling ASD
[b]Mullen Scales of Early Learning, Age Equivalent
[c]Vineland Adaptive Behavior Scales
[d]Autism Diagnostic Observation Schedule
[e]Social Affect Score
[f]Restricted and Repetitive Behaviors Score

group based on a diagnosis given between 18 and 30 months (mean age 24.2 months).

Movie, apparatus, and eye tracking procedure

The Complex Social GeoPref test contained a 90-s movie composed of two large, rectangular areas of interest (AOIs) side by side (see Fig. 1) where one AOI displayed geometric patterns and the other social scenes. There was no audio. The geometric patterns were a subset of those of the original 60-s GeoPref test [38]; however, each geometric pattern was repeated for a longer time interval to achieve a 90-s test. Each social scene was paired with one of the moving, colorful geometric patterns, and when each social scene changed, each geometric pattern also changed. The social scenes included five scenes showing two children interacting. The interactions were dance-like twisting side by side, jumping to a high-five, whispering a secret then appearing surprised, whispering a secret then hugging, and teasing by sticking out tongues then stomping on the other's foot. To allow for cross-paradigm comparisons, a subset of 126 toddlers participated in the original GeoPref test identical to Pierce et al. [38] as well as the new Complex Social GeoPref test on separate visits. Unlike the complex social scenes in the present design, in the original GeoPref test, all scenes similarly showed children doing rhythmic, dance-like movements all displaying a uniformly positive affect. See Fig. 1 for sample images and the movie clip "Additional file 2" for more details.

Eye tracking data were collected while toddlers, seated on a parent's lap, viewed these videos from 60 cm distance on a 17″ thin-film transistor monitor using the Tobii T-120 system set at 60 Hz. Five-point calibration was first performed with Tobii Studio software using an animated image with sound presented at known *X*-*Y* coordinates. Eye tracking data were collected only if the calibration result fell within the parameters reported by the manufacturer to yield an accuracy of 0.5° [49]. Each AOI subtended 12.9° horizontally and 9.1° vertically. This use of large, simple AOIs facilitated correct measurement of the infant/toddler population, who can yield data with accuracy below levels reported for adults under optimal conditions [50]. More information about data spatial accuracy is provided in the Additional file 1.

Fixations were classified based on gaze data averaged from both eyes using a velocity threshold Tobii Fixation

Filter set to 35 pixels/window, which interpolates to fill in data loss of less than 100 ms. For each subject, number of fixations, duration of each fixation, and sum of fixation time within the two AOIs (social and geometric) were calculated. Sum of fixation time per AOI was divided by total sum of fixation time for both AOIs to derive proportion of time spent on each AOI (i.e., "%Geo" and "%Soc") and to correct for missing data. Subjects with excessive missing data (i.e., less than 30 s of data) due to attending to neither AOI or due to inability to track eye gaze (e.g., during excessive movement) were excluded, in order to preclude inaccurate measurement of number or length of fixations and saccades. Number of fixations per AOI was divided by sum of fixation time for that AOI to derive saccade frequency as saccades per second, which was also reported in our prior publication [39]. See the Additional file 1: Figure S1 for the complete description of exclusion criteria and lab practices for assuring data accuracy and precision, and also the results regarding saccade per second differences between groups.

Statistical analyses

Percent of total fixation duration to geometric vs complex social stimuli

No age differences were found between diagnostic groups (one-way ANOVA with no overall effect of age; see Table 1). To compare percentage of total fixation time within the geometric AOI between groups, a one-way ANOVA was performed (diagnostic group (6 levels) × %Geo (1 level)). A significant main effect was followed by Bonferroni-corrected post hoc pairwise comparisons. Prior to selecting these analysis strategies, homogeneity of variance was confirmed with Levene's test. To confirm that differences in data quality between diagnostic groups were not impacting the reported results, an ANCOVA was performed as well, with six diagnostic groups as a fixed factor, %Geo as the dependent variable, and a data quality measure (percent of valid samples obtained) as a covariate. There was no significant effect of the data quality metric ($F_{5,221} = .011$, $p = .916$).

Relationships between percent of total fixation duration and clinical characteristics

All statistical values for clinical scores presented in Table 1, namely, those of ADOS, MSEL, and VABS, were calculated in the same manner: one-way ANOVAs were performed with post hoc pairwise tests and Bonferroni corrections. The relationship between Complex Social GeoPref %Geo scores and ADOS total scores was assessed with Spearman's rank-order correlations. After identifying the ASD children with the strongest preferences for geometric and for social images using %Geo and %Soc scores, within ASD analyses focused on differences in clinical characteristics between the subtypes "GeoPref" and "SocPref" followed strategies like those described above: homogeneity of variance was assessed and no significant age difference between groups was found, so independent samples *t* tests were used to compare scores on ADOS, MSEL, and VABS scores. These comparisons presented in Table 2 are reported one-tailed based on a priori hypotheses from our 2016 manuscript regarding the direction of differences. Correction for multiple comparisons was performed using the Benjamini-Hochberg procedure. Cohen's *d* effect sizes are reported as well.

Clinical classification performance: sensitivity, specificity, PPV, NPV, and ROC curve

To assess the ability of the Complex Social GeoPref test to discriminate toddlers with ASD from other toddlers, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and Receiver Operating Characteristic (ROC) area under the curve were determined. For consistency and comparison with our past publications, 69% looking time to geometric images was used as the cut-off for a positive result. Although PPV and NPV for a general population ASD screening tool would be calculated based on the 1/68 prevalence rate for ASD [14], the GeoPref tests are best suited as second

Table 2 Participant characteristics of ASD subgroups

	ASD GeoPref subtype	ASD SocPref subtype	*t* test, corrected *p*	Cohen's *d*
Sex, M/F	13/1	15/2	n/a	
Age at eye tracking, months Mean (SD) [range]	31.2 (9.2) [13.8–47.2]	28.7 (7.2) [16.4–45.4]	.40	.30
MSEL[a] age equivalent scores/true age				
Visual reception	.76 (.17)	.81 (.23)	.50	.25
Fine motor	.76 (.19)	.82 (.18)	.41	.32
Receptive language	.46 (.25)	.68 (.26)	.03	.86
Expressive language	.53 (.26)	.67 (.30)	.19	.50
VABS[b] standard scores				
Communication	71.0 (16.0)	81.5 (16.0)	.08	.66
Daily living	80.9 (15.2)	84.1 (15.3)	.57	.21
Socialization	77.1 (12.7)	84.8 (14.4)	.13	.57
Motor skills	84.4 (10.7)	95.7 (15.0)	.03	.87
Composite score	75.6 (12.7)	84.2 (15.9)	.11	.60
ADOS[c] module T, 1, or 2				
SA[d]	16.7 (3.2)	11.4 (4.1)	< .001	1.4
RRB[e]	4.2 (2.5)	3.1 (1.7)	.15	.51
Total score	20.9 (4.4)	14.5 (4.2)	< .001	1.5

[a]Mullen Scales of Early Learning
[b]Vineland Adaptive Behavior Scales
[c]Autism Diagnostic Observation Schedule
[d]Social affect
[e]Restricted and repetitive behaviors

tier tools, administered after a questionnaire screener which has higher sensitivity but lower specificity. Therefore, PPV and NPV were calculated here against the ASD rate in our sample (i.e., 1/3). This rate reflects a PPV and NPV that might be expected at a general ASD and developmental disorder diagnosis and evaluation clinic, where children are referred primarily due to failing a first-tier screening tool (i.e., the CSBS-DP Infant Toddler Checklist). Classification statistics are presented separately for the entire sample and for screening referred children only without including self-referred children. However, it is to be expected that in a real-world clinical setting, self-referrals will naturally occur as there are many ways outside of pediatrician screening (e.g., Google searching) that community members might become aware of the availability of evaluation services and then self-refer.

Because the greatest challenge to clinicians is distinguishing ASD toddlers from toddlers with other sorts of delays, these classification performance measures were also calculated without the inclusion of TD and TypSib groups and are presented in the Additional file 1. Because 69% was chosen in our previous work by setting the test's specificity to 99%, classification performance values are also reported for the cut-off that gives a specificity of 99% on the Complex Social GeoPref test (75% of looking time to geometric images) in the Additional file 1. Use of the Complex Social GeoPref test in order to rule out a diagnosis of ASD, where having a %Geo score below a certain threshold is considered positive, plus having any diagnosis other than ASD is considered true positive, is examined in the Additional file 1 with regard to sensitivity, specificity, PPV, NPV, and AUC.

Comparing and combining of complex social and original GeoPref tests

Differences between the Complex Social GeoPref and original GeoPref tests for the subset of children who completed both tests in the percentage of time viewing geometric stimuli (%Geo scores) were investigated in several ways. Paired samples *t* tests were used to compare %Geo scores for the two tests for each diagnostic group. Degree of correlation between test scores for individual children who completed both tests was assessed with Spearman's rank-order correlation. Use of both tests by a single child, where a positive score on either test (or both tests) is considered a positive result, was also examined with regard to sensitivity, specificity, PPV, NPV, and AUC. AUC for this two-test model was determined based on predicted probabilities calculated using binary logistic regression with %Geo scores for the two tests as covariates.

Results

Percent of total fixation duration of the six diagnostic groups to geometric vs complex social stimuli

In our new Complex Social GeoPref test, geometric images attracted significantly more looking time in ASD than in TD, DD, and other groups ($F_{5,221} = 9.1$, $p < .001$, partial eta-squared = .17; ASD vs TD, $p < .001$, Cohen's $d = .85$; ASD vs DD, $p < .001$, Cohen's $d = 1.0$; ASD vs other, $p < .005$, Cohen's $d = .96$). Toddlers with ASD spent an average of 48.4% of their time looking at geometric images (95% confidence interval (CI) range 43.6–53.2%); TD toddlers 31.2% of their time (95% CI 25.8–36.6%); DD toddlers 28.6% of their time (95% CI 23.8–33.4%); and other toddlers 30.0% of their time (95% CI 22.5–37.6%). Toddlers with ASD also looked more at geometric images than TypSibs (mean 32.8%, 95% CI 20.8–44.7%) and ASD features (mean 39.0%, 95% CI 28.9–49.0%), but these differences were not statistically significant, perhaps due to small sample sizes in the latter two study groups. See Fig. 2.

Within ASD: differences between the ASD GeoPref and ASD SocPref subtypes

Percent of total fixation duration per AOI and clinical characteristic comparisons

In order to compare clinical characteristics associated with ASD toddlers at either end of the fixation spectrum [39] (i.e., those who strongly preferred geometric images and those who strongly preferred social images), ASD toddlers were identified who were either the ASD GeoPref subtype (> 69% time looking at geometric images) or the ASD SocPref subtype (> 69% time looking at complex social images). The mean %Geo looking score for the ASD GeoPref subtype was 80.5% (or 19.5% social looking), and the mean %Soc looking score for the ASD SocPref subtype was 80.3% (or 19.7% geometric looking).

Clinical differences between the two ASD subtypes are shown in Table 2. Similar to our previous reports [39], as compared to the ASD SocPref subjects, the ASD GeoPref subjects had significantly increased ADOS social affect ($t_{29} = 4.0$, $p < .001$, Cohen's $d = 1.4$) and total scores ($t_{29} = 4.2$, $p < .001$, Cohen's $d = 1.5$). Further, across the entire group of 76 toddlers with ASD, %Geo scores on the Complex Social GeoPref test were significantly correlated with ADOS total scores, that is, those with greater preference for looking at the geometric stimuli had more severe autism scores (Spearman's rho = .43, $p < .001$), and this is presented in the Additional file 1. The ASD GeoPref subjects appeared to also have lower mean Mullen receptive language scores, and lower Vineland motor scores, with moderate Cohen's *d* effect sizes.

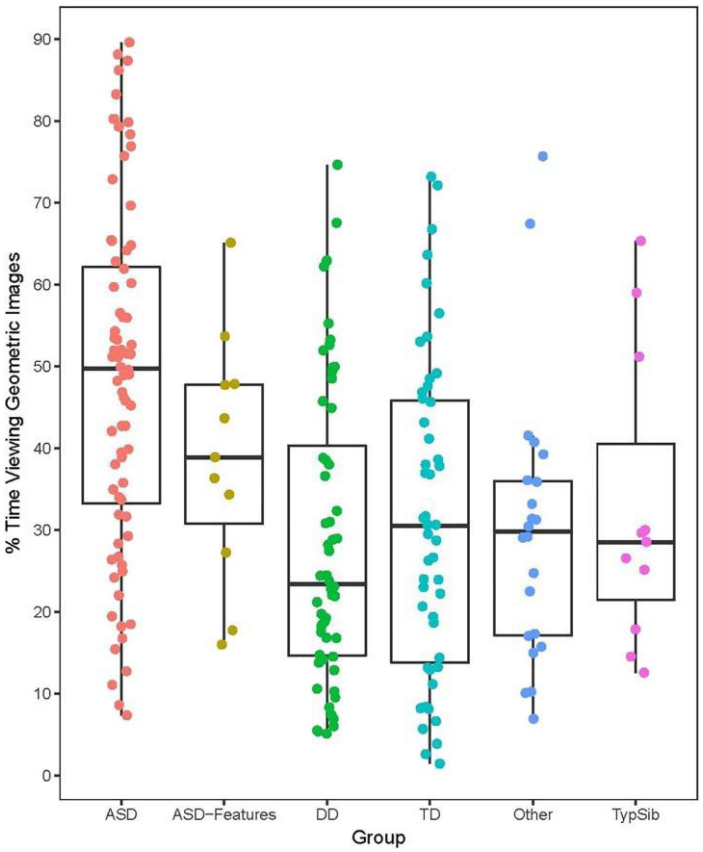

Fig. 2 Scatterplot of % time viewing geometric images (total fixation duration viewing the dynamic geometric stimulus, divided by total fixation duration to the geometric and social stimuli combined) for all subjects who completed the Complex Social GeoPref test (n = 227) sorted by diagnostic group. Boxplots show median, range, and first and third quartiles

New Complex Social GeoPref test vs original GeoPref test

Eye tracking data were examined from 126 of the 227 study subjects (37 ASD, 10 ASD features, 30 DD, 32 TD, 7 Typ-Sibs, 10 other) who completed both the new Complex Social GeoPref test and the original GeoPref test on separate visits to the center. Sixty-seven completed the original GeoPref test first (mean age 27.3 months at original GeoPref testing) and 59 completed the Complex Social GeoPref test first (mean age 29.0 months at Complex Social GeoPref testing), and this age difference was not significant. Further, no diagnostic group differences in age at testing were found.

Across all diagnostic groups, the mean difference between the Complex Social GeoPref and original GeoPref tests in percent time looking at geometric images (%Geo) was 1.3% (%Geo = 35.4% for the Complex Social vs %Geo = 34.1% for the original GeoPref test), and this was not significant (t_{125} = 0.6, p = .60). For the ASD group, the mean %Geo score was 48.7% for the Complex Social and 44.7% for the original GeoPref tests, which was a non-significant difference (t_{35} = 1.0, p = .31).

There was a significant within-subject correlation in %Geo of total fixation duration across the Complex Social and original GeoPref tests across all study subjects (N = 126; Spearman's r = .25, p < .005) and within the ASD group (N = 37; Spearman's r = .47, p < .005). See Fig. 3. For TD and DD groups, these scores were not significantly correlated.

ROC comparisons were all examined, with rates of specificity, sensitivity, PPV, NPV, and AUC compared between the Complex Social and original GeoPref tests in Table 3. AUC was 0.74 for the Complex Social GeoPref test. For comparison, AUC was 0.71 for the original GeoPref test in Pierce et al. (2016). Removal of typically developing children from analysis and focusing on children with some sort of delay or disorder further improved AUC to 0.75 (see Additional file 1). Re- moval of self-referred children and focusing on screening referred only yielded an AUC of 0.73.

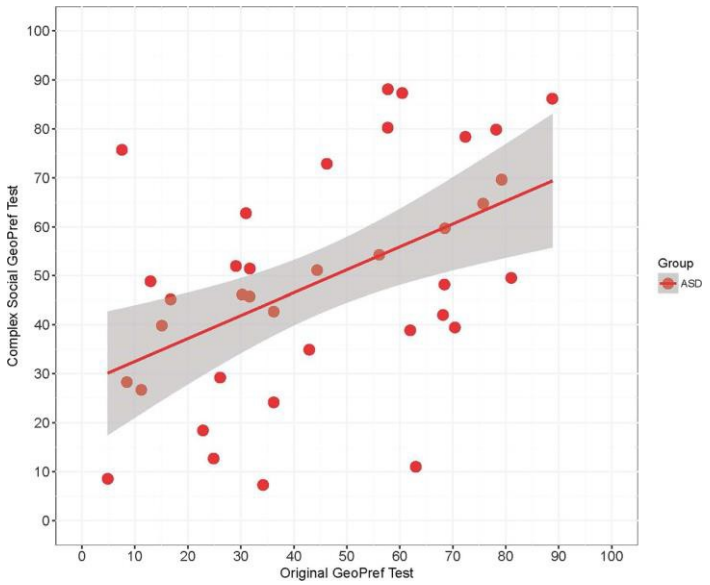

Fig. 3 Correlation between original and Complex Social GeoPref Tests. Scatterplot illustrating % Geo scores (summed fixation duration while viewing the dynamic geometric stimulus, divided by total fixation duration to the geometric and social stimuli combined) for each test for each subject in the ASD group who completed both eye tracking tests on separate visits across the span of the study

Clinical classification performance: combination of new and original GeoPref tests

AUC, sensitivity, specificity, and PPV and NPV values were then examined when using a two-test screening model on data from the 126 subjects who participated in both the Complex Social and original GeoPref tests in order to determine if use of two tests enhanced classification performance. Results are shown in Table 4 (see data in the left two columns). Sensitivity increased substantially from 18% for the Complex Social test alone (Table 3) to 35% with two tests, while specificity remained high. AUC calculated for this two-test model was 0.75. Pilot results in the Additional file 1 suggest this enhancement does not occur if the two tests are given immediately back-to-back but only if separated in time.

Table 3 Clinical classification performance, comparison of original and Complex Social Geopref tests

69% Geo threshold	All available subjects *N* = 444	All available subjects *N* = 227	Screening referred subjects only *N* = 122
True positive = ASD only	Original GeoPref test (2016)	Complex Social GeoPref test	Complex Social GeoPref test
True positive	35	14	10
False negative	117	62	36
False positive	4	4	3
True negative	288	147	73
Sensitivity (%)	23	18	22
Specificity (%)	99	97	96
Positive predictive value (%)	90	78	77
Negative predictive value (%)	71	70	67
Area under ROC curve	.71	.74	.73

Classification performance of the combination of the two tests for screening referred toddlers only was also examined. Results are shown in Table 4 (data in the right two columns). Again, sensitivity increased substantially from 18% (Table 3) to 33%, while specificity remained high. AUC calculated for this two-test model applied to screening referred toddlers only was 0.73.

Discussion

Debate and controversy regarding the replication of findings from the biological sciences and psychology have been common in recent years [51, 52]. In contrast, the Geometric Preference effect in toddlers with autism has now been replicated multiple times in both direct, identical replication [39] and in this conceptual replication with varied social stimuli. Following our original report in 2011 [38], independent laboratories have also reported similar findings, e.g., [53, 54]. The Complex Social GeoPref test has 97% specificity for ASD in our sample, which is especially high given that our sample contains toddlers with a large variety of presentations beyond typical development and ASD, including

Table 4 Clinical classification, two tests: combined original and Complex Social Geopref Tests

69% Geo threshold	All available subjects *N* = 126	All available subjects *N* = 126	Screening referred subjects only *N* = 82	Screening referred subjects only *N* = 82
Positive = positive on either test	ASD only = true positive	ASD + ASD features = true positive	ASD only = true positive	ASD + ASD features = true positive
True positive	13	15	9	11
False negative	24	32	18	23
False positive	5	3	5	3
True negative	84	76	50	45
Sensitivity (%)	35	32	33	32
Specificity (%)	94	96	91	94
Positive predictive value (%)	72	83	64	79
Negative predictive value (%)	78	70	74	66
Area under ROC curve	.75	.78	.73	.76

language delay and global developmental delay. In comparison, genetic biomarkers of ASD are often pleiotropic and therefore also associated with a number of other neurodevelopmental disorders, so they can have poor specificity [1], as well as low sensitivity due to the large number of different genetic inputs that converge on the ASD phenotype [55]. Effective usage of GeoPref tests would involve prescreening such as we have done here using the CSBS at pediatrician offices; therefore, positive predictive value need not be measured against the 1/68 base rate of ASD in the general population [9]. Without high specificity tests, applied correctly, with results communicated appropriately, false positives do result in inadvertent harms in the process of early identification for infants at risk for ASD [56]. These harms include the family's exposure to stress and stigma and the unnecessary usage of somewhat scarce and costly intervention services [57]. It has been shown that pediatricians do not refer a significant portion of children who fail screenings for developmental delays, probably due in part to concern regarding the potential for false positive results [58]. Therefore, the availability of screening tools with few false positives could significantly impact the efficacy of screening procedures used for early identification of ASD.

Because the GeoPref tests, both the original and the Complex Social version, detect a subtype of ASD, sensitivity is considerably lower than specificity: at optimal specificity, the Complex Social GeoPref test will catch about 1 in 5 children with autism, while the rate for the original GeoPref test is about 1 in 4.However, here we show that when the two tests are used in combination across separate testing sessions, the correct detection rate is 1 in 3 and specificity remains high at about 94% (Table 4). Further, as can be seen in Fig. 2, the range of %Geo scores for ASD children does not extend as low as that of other groups, which is not a property of the original GeoPref test. If borne out by further data, this could be of value clinically as a means of ruling out ASD in certain children who are exhibiting some ambiguous warning signs but have very low %Geo scores, in order to shift them away from unnecessary ASD services. This result is described in more detail in the Additional file 1. Future research will work toward creating a battery of multiple eye tracking tests in order to further increase sensitivity to ASD in general and to zero in on optimal procedures for detection of this GeoPref subtype of ASD and to elucidate its biological bases.

The ASD GeoPref subtype toddlers detected with GeoPref tests tend to be the most affected cases, as ADOS symptom severity is correlated with %Geo score. It has been observed that more severe presentations of ASD tend to be less studied [59], despite being arguably more in need of treatment. It is possible that the defects impacting the "social brain," particularly in the frontal regions that control attention and social interest [60], are more pronounced in this ASD subgroup. Since functional brain imaging began to be utilized to understand the operations of the brains of those on the autism spectrum about 20 years ago, abnormalities in virtually every social brain region examined have been reported [61]. However, in addition to the fact that such studies almost exclusively included only older and/or high functioning individuals, data was almost always presented at the group level. As such, previous studies made it hard to understand if reported "social brain" abnormalities were ubiquitous across all ASD individuals, or were being driven by certain subgroups or individuals with the most severe functional abnormalities. In an effort to parse the heterogeneity of social brain neural functional responding in ASD, our new resting state functional imaging study, which examined ASD GeoPref toddlers as a separate subgroup, found substantially weakened functional connectivity between the default mode network

(DMN) which includes key "social brain" regions such as the medial prefrontal cortex [62] and a visual network within the occipito-temporal cortex (OTC) in GeoPref ASD toddlers, but not in other toddlers (Lombardo et al., In Review). This finding is consistent with the previous theory that argues that ASD is a disorder wherein higher order social frontal systems are disconnected from more basic systems [63] and further underscores that the severity of this disconnection may be a driving factor in the social abilities of ASD individuals. Notably, ASD toddlers that did not show the GeoPref profile, i.e., SocPref ASD toddlers, did not show distinctly abnormal functional connectivity between the DMN and OTC (Lombardo et al., In Review). The implication of this work is that ASD SocPref toddlers may have stronger and more typical functional circuitry and, again, promise for a better long-term outcome.

Given the intrinsic heterogeneity in the loose category circumscribed by an ASD diagnosis, focusing on subgroups with phenotypic commonalities may be a key research strategy [64]. Another topic for future research is characterization of the GeoPref and SocPref subtypes in terms of traits that are prevalent but not defining characteristics of ASD, such as gastrointestinal issues, altered sensorimotor processing, or comorbid seizure disorder. If found, differences in rates of comorbid epilepsy, motor impairment, and sleep disturbance, because specific mutations have been associated with each [65], could point to genotypic differences between the phenotypic subgroups identified by GeoPref tests.

Consistent with previously reported findings [43], we hypothesized that our revised social stimuli that presented more than one person or social interactions between multiple people tend to magnify the differences between ASD and TD gaze behavior when compared to simpler social stimuli presenting a single person. We did not find this to be the case, as the Original GeoPref test that paired individual children dancing and dynamic geometric images elicited similar or even slightly larger differences between diagnostic groups than the current Complex Social GeoPref test. Alternately, other variables that differ between the Original and Complex Social stimuli, such as salience of biological motion, temporal dynamics of vignettes unfolding, or the overall length, or perhaps differences in low-level visual properties influencing salience (e.g., color or contrast), may account for this finding [66, 67].

One limitation of the current study is that because of differences between the stimuli that are unrelated to their content as social and geometric (e.g., basic feature salience), "geometric preference" is not the only reasonable explanation for the observed differences in behavior across groups. Although we have referred to ASD children who show the least interest in the social stimuli and the most interest in dynamic geometric images as "Geometric Responders" or the "GeoPref" subtype, we have not yet manipulated the geometric images in a large study. It is conceivable that pronounced lack of interest in, or aversion to, the social stimuli alone is driving the geometric preferences, and one could replace the competing stimuli with another type of nonsocial stimuli with similar results. At least one new study suggests that aversion to gaze is not a driving factor in abnormal visual fixation patterns in ASD [68]. However, atypical amygdala responses when viewing eye gaze and faces and disrupted amygdala functional connectivity have been previously observed in ASD and related to gaze aversion and social anxiety [69–71]. It is also possible that the slow rate of saccades shown by the GeoPref subgroups (see Additional file 1) to geometric stimuli may indicate difficulty with attentional disengagement, which then causes longer percent total fixation duration to geometric images. This explanation may have little to do with social motivation, social reward, or "preference" for one stimulus type over the other. Interestingly, however, research studies examining attentional disengagement in ASD are inconclusive, with reports of both deficits [21] and typical responding [72], likely reflecting the wide range of stimuli and procedures used across studies.

The importance of finding the GeoPref profile in toddlers may go beyond its potential value as a screening or even diagnostic biomarker—it may be most importantly useful as a prognostic biomarker. Although it is currently unknown if the abnormal visual fixation patterns displayed during the Complex Social and original GeoPref tests generalize to the everyday life of ASD toddlers, it is at least theoretically plausible that ASD toddlers who display the GeoPref profile are experiencing socially impoverished visual input from their environment. As experience in the first few years of life crucially shapes the brain's organization, we hypothesize that a GeoPref profile in a toddler may predict distal functional and cognitive outcomes, and our future work intends to examine whether the GeoPref profile is associated with a worse long-term outcome than that of ASD toddlers who prefer social images. Importantly, experience-dependent mechanisms involved in early social learning may be amenable to intervention, and therefore, GeoPref tests may be useful for early identification of and differential intervention for toddlers who strongly attend to certain non-social stimuli, ignoring social information. Tailoring treatment according to ASD subtypes could potentially result in improved treatment responses and better long-term outcomes.

Conclusion

Across multiple types of social stimuli and temporal presentations, substantially increased viewing of geometric

images during preferential looking tasks that pair dynamic social and geometric images robustly indicates ASD among toddlers. Furthermore, across multiple sorts of stimuli, the subset of ASD toddlers who strongly prefer geometric images have more severe scores on indicators of autism impairment compared to those who strongly prefer social images. In addition to replicating the original GeoPref phenomenon, the Complex Social GeoPref test finding shows potential as a valid behavioral biomarker, as it identifies ASD in toddlers at the individual subject level.

Additional files

Additional file 1: Supplementary Text. (PDF 459 kb)
Additional file 2: Stimulus Video Example. (MP4 14305 kb)

Acknowledgements
We thank Sunny Pence, Mee-Kyoung Kwon, and Linda Lopez for the assistance with the data collection, subject tracking, and research assistant supervision, Srinivasa Nalabolu for the data management, Elizabeth Bacon and Jessica Garcia for the assistance with the clinical testing, Josiah Faustino for the data quality confirmation, and Vahid Gazestani for the statistical consulting.

Funding
This work was supported by the National Institutes of Health grants R01-MH080134 and P50-MH081755.

Availability of data and materials
The datasets analyzed during the current study will be made available through the National Database for Autism Research (NDAR) repository in 2017. Currently, the datasets are available from the corresponding author on reasonable request.

Authors' contributions
AM collected the data, analyzed the data, and wrote and revised the manuscript. MW and AY collected the data and contributed to the data analysis. CCB and DC provided the diagnostic and standardized testing for the participants. EC interpreted the findings and revised the manuscript. KP designed the study, interpreted the findings, and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
This study was approved by the University of California, San Diego, Human Subjects Research Protection Program (IRB #081722). Legal guardians of all participants gave written informed consent.

Consent for publication
Written informed consent was received for publication of their images in Fig. 1b and Additional file 1: Figure S1 and Additional file 2 of this manuscript from parents/legal guardians of the children. The consent forms are held by the authors and are available for review by the Editor-in-Chief.

Competing interests
An invention disclosure form was filed by KP with the University of California, San Diego, on March 5, 2010, and the original GeoPref test is licensed by the University of California, San Diego, but free for research use. The other authors declare that they have no competing interests.

Publisher's Note

Received: 11 April 2017 Accepted: 2 March 2018
Published online: 21 March 2018

References
1. Walsh P, et al. In search of biomarkers for autism: scientific, social and ethical challenges. Nat Rev Neurosci. 2011;12(10):603–12.
2. Tager-Flusberg H, Kasari C. Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. Autism Res. 2013;6(6):468–78.
3. Bal VH, et al. Daily living skills in individuals with autism spectrum disorder from 2 to 21 years of age. Autism. 2015;19(7):774–84.
4. Newschaffer CJ, Curran LK. Autism: an emerging public health problem. Public Health Rep. 2003;118(5):393–9.
5. Dawson G, et al. Early behavioral intervention is associated with normalized brain activity in young children with autism. J Am Acad Child Adolesc Psychiatry. 2012;51(11):1150–9.
6. Dawson G, et al. Randomized, controlled trial of an intervention for toddlers with autism: the early start Denver model. Pediatrics. 2010;125(1):e17–23.
7. Anderson DK, Liang JW, Lord C. Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. J Child Psychol Psychiatry. 2014;55(5):485–94.
8. Bauman ML, Kemper TL. Neuroanatomic observations of the brain in autism: a review and future directions. Int J Dev Neurosci. 2005;23(2–3):183–7.
9. Anderson GM. Autism biomarkers: challenges, pitfalls and possibilities. J Autism Dev Disord. 2015;45(4):1103–13.
10. Stoner R, et al. Patches of disorganization in the neocortex of children with autism. N Engl J Med. 2014;370(13):1209–19.
11. Shic F, Macari S, Chawarska K. Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. Biol Psychiatry. 2014;75(3):231–7.
12. Jones W, Klin A. Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. Nature. 2013;504(7480):427–31.
13. Chawarska K, Macari S, Shic F. Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. Biol Psychiatry. 2013;74(3):195–203.
14. Christensen DL, Baio J, Van Naarden Braun K, Bilder D, Charles J, Constantino JN, Daniels J, Durkin MS, Fitzgerald RT, Kurzius-Spencer M, Lee LC, Pettygrove S, Robinson C, Schulz E, Wells C, Wingate MS, Zahorodny W, Yeargin-Allsopp M. Prevalence and characteristics of autism spectrum disorder among children aged 8 years–autism and developmental disabilities monitoring network, 11 sites, United States, 2012. Centers for Disease Control and Prevention (CDC): MMWR. Surveillance Summaries; 2016.
15. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol Psychiatry. 2012;17(12):1174–9.
16. Insel TR. The NIMH Research Domain Criteria (RDoC) project: precision medicine for psychiatry. Am J Psychiatry. 2014;171(4):395–7.
17. Bolte S, et al. How can clinicians detect and treat autism early? Methodological trends of technology use in research. Acta Paediatr. 2016;105(2):137–44.
18. Gredebäck G, Johnson S, von Hofsten C. Eye tracking in infancy research. Dev Neuropsychol. 2010;35(1):1–19.
19. Rommelse NN, Van der Stigchel S, Sergeant JA. A review on eye movement studies in childhood and adolescent psychiatry. Brain Cogn. 2008;68(3):391–414.
20. Wass SV, et al. Shorter spontaneous fixation durations in infants with later emerging autism. Sci Rep. 2015;5:8284.
21. Sacrey LA, et al. Impairments to visual disengagement in autism spectrum disorder: a review of experimental studies from infancy to adulthood. Neurosci Biobehav Rev. 2014;47:559–77.
22. Falck-Ytter T, Bolte S, Gredeback G. Eye tracking in early autism research. J Neurodev Disord. 2013;5(1):28.
23. Insel TR. Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. Arch Gen Psychiatry. 2009;66(2):128–33.
24. Klin A, et al. Two-year-olds with autism orient to non-social contingencies rather than biological motion. Nature. 2009;459(7244):257–61.
25. Shic F, et al. Limited activity monitoring in toddlers with autism spectrum disorder. Brain Res. 2011;1380:246–54.
26. Chawarska K, Shic F. Looking but not seeing: atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. J Autism Dev Disord. 2009;39(12):1663–72.

27. Thorup E, et al. Altered gaze following during live interaction in infants at risk for autism: an eye tracking study. Mol Autism. 2016;7:12.
28. Bedford R, et al. Precursors to social and communication difficulties in infants at-risk for autism: gaze following and attentional engagement. J Autism Dev Disord. 2012;42(10):2208–18.
29. Carpenter M, Nagell K, Tomasello M. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monogr Soc Res Child Dev. 1998;63(4):i–vi. 1-143
30. Charman T. Why is joint attention a pivotal skill in autism? Philos Trans R Soc Lond Ser B Biol Sci. 2003;358(1430):315–24.
31. Jones W, Carr K, Klin A. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. Arch Gen Psychiatry. 2008;65(8):946–54.
32. Keehn B, Muller RA, Townsend J. Atypical attentional networks and the emergence of autism. Neurosci Biobehav Rev. 2013;37(2):164–83.
33. Elsabbagh M, et al. Social and attention factors during infancy and the later emergence of autism characteristics. Prog Brain Res. 2011;189:195–207.
34. Bedford R, et al. Additive effects of social and non-social attention during infancy relate to later autism spectrum disorder. Dev Sci. 2014;17(4):612–20.
35. Murza KA, et al. Joint attention interventions for children with autism spectrum disorder: a systematic review and meta-analysis. Int J Lang Commun Disord. 2016;51(3):236–51.
36. Chawarska K, Macari S, Shic F. Context modulates attention to social scenes in toddlers with autism. J Child Psychol Psychiatry. 2012;53(8):903–13.
37. Guillon Q, et al. Visual social attention in autism spectrum disorder: insights from eye tracking studies. Neurosci Biobehav Rev. 2014;42:279–97.
38. Pierce K, et al. Preference for geometric patterns early in life as a risk factor for autism. Arch Gen Psychiatry. 2011;68(1):101–9.
39. Pierce K, et al. Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. Biol Psychiatry. 2016;79(8):657–66.
40. Hadjikhani N, et al. Abnormal activation of the social brain during face perception in autism. Hum Brain Mapp. 2007;28(5):441–9.
41. Grossman ED, Blake R. Brain areas active during visual perception of biological motion. Neuron. 2002;35(6):1167–75.
42. Lloyd-Fox S, et al. Selective cortical mapping of biological motion processing in young infants. J Cogn Neurosci. 2011;23(9):2521–32.
43. Chita-Tegmark M. Social attention in ASD: a review and meta-analysis of eye-tracking studies. Res Dev Disabil. 2016;48:79–93.
44. Pierce K, et al. Detecting, studying, and treating autism early: the one-year well-baby check-up approach. J Pediatr. 2011;159(3):458–65. e1-6
45. Wetherby AM, et al. Validation of the Infant-Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. Autism. 2008;12(5):487–511.
46. Luyster R, et al. The Autism Diagnostic Observation Schedule-Toddler Module: a new module of a standardized diagnostic measure for autism spectrum disorders. J Autism Dev Disord. 2009;39(9):1305–20.
47. Mullen E. Mullen Scales of Early Learning. Circle Pines: American Guidance Service; 1995.
48. Sparrow S. *Vineland Adaptive Behavior Scales: Survey Form Manual*, D. Balla, Editor. Circle Pines, MN: American Guidance Service; 1984.
49. Accuracy and precision test method for remote eye trackers. 2011 [cited Accessed 18 Feb 2017.
50. Wass SV, Smith TJ, Johnson MH. Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. Behav Res Methods. 2013;45(1):229–50.
51. Pashler H, Wagenmakers EJ. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? Perspect Psychol Sci. 2012;7(6):528–30.
52. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. Nature. 2012;483(7391):531–3.
53. Franchini M, et al. Brief report: a preference for biological motion predicts a reduction in symptom severity 1 year later in preschoolers with autism spectrum disorders. Front Psychiatry. 2016;7:143.
54. Franchini M, et al. Social orienting and joint attention in preschoolers with autism spectrum disorders. PLoS One. 2017;12(6):e0178859.
55. Goldani AA, et al. Biomarkers in autism. Front Psychiatry. 2014;5:100.
56. Rossi J, Newschaffer C, Yudell M. Autism spectrum disorders, risk communication, and the problem of inadvertent harm. Kennedy Inst Ethics J. 2013;23(2):105–38.
57. Sheldrick RC, et al. Thresholds and accuracy in screening tools for early detection of psychopathology. J Child Psychol Psychiatry. 2015;56(9):936–48.
58. Wissow LS, et al. Universal mental health screening in pediatric primary care: a systematic review. J Am Acad Child Adolesc Psychiatry. 2013;52(11): 1134–47. e23
59. Jack A, Pelphrey KA. Annual research review: understudied populations within the autism spectrum—current trends and future directions in neuroimaging research. J Child Psychol Psychiatry. 2017;58(4):411–35.
60. Bicks LK, et al. Prefrontal cortex and social cognition in mouse and man. Front Psychol. 2015;6:1805.
61. Adolphs R. Cognitive neuroscience of human social behaviour. Nat Rev Neurosci. 2003;4(3):165–78.
62. Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci. 2006;7(4):268–77.
63. Courchesne E, Pierce K. Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. Curr Opin Neurobiol. 2005;15(2):225–30.
64. Voineagu I, Yoo HJ. Current progress and challenges in the search for autism biomarkers. Dis Markers. 2013;35(1):55–65.
65. Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. Nat Rev Neurol. 2014;10(2):74–81.
66. Bertone A, et al. Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. Brain. 2005;128(Pt 10):2430–41.
67. Perreault A, et al. Behavioral evidence for a functional link between low- and mid-level visual perception in the autism spectrum. Neuropsychologia. 2015;77:380–6.
68. Moriuchi JM, Klin A, Jones W. Mechanisms of diminished attention to eyes in autism. Am J Psychiatry. 2017;174(1):26–35.
69. Shen MD, et al. Functional connectivity of the amygdala is disrupted in preschool-aged children with autism spectrum disorder. J Am Acad Child Adolesc Psychiatry. 2016;55(9):817–24.
70. Kleinhans NM, et al. Association between amygdala response to emotional faces and social anxiety in autism spectrum disorders. Neuropsychologia. 2010;48(12):3665–70.
71. Baron-Cohen S, et al. The amygdala theory of autism. Neurosci Biobehav Rev. 2000;24(3):355–64.
72. Fischer J, et al. Unimpaired attentional disengagement in toddlers with autism spectrum disorder. Dev Sci. 2016;19(6):1095–103.

Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in Molecular Autism, 2018. Moore, Adrienne; Wozniak, Madeline; Yousef, Andrew; Barnes, Cindy Carter; Cha, Debra; Courchesne, Eric; Pierce, Karen. Adrienne Moore was the primary author and investigator of this paper.

Additional file 1

Additional sample information

Eye tracking data from the Complex Social GeoPref test were collected from 2013 to 2015. If participants were eye tracked multiple times across longitudinal visits, data included are from the earliest appointment available. If clinical assessments were completed multiple times across longitudinal visits, scores included are from the same day when Complex Social GeoPref data were collected whenever possible. Details of the reasons for subject exclusion and number excluded per diagnosis are summarized in Figure S1 below.

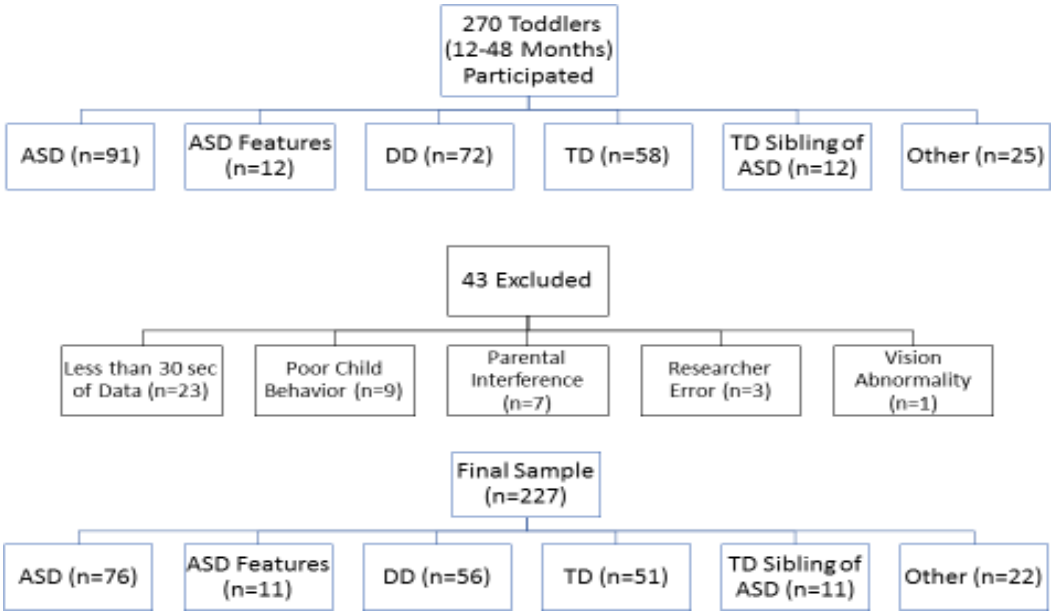

Figure 3.S1: Summary of subject inclusion and exclusion. Note: Excluded refers to children for whom eye tracking data was obtained, but was unusable. 4 additional toddlers (2 DD, 1 ASD, 1 Other) would have participated, bringing the total participants to 274, however researchers were unable to obtain any eye tracking data due to inability to calibrate the eye tracker to the child's eye.

Additional data quality information

Eye tracking data collection typically occurred first or early in an evaluation session, prior to other assessments. Since 2015 we have used spatial accuracy verification testing consisting of a looming cross paired with sound subtending 3.5 degrees at its largest size presented at three positions (top left, top right, and center) immediately after testing to determine whether excessive loss of spatial accuracy has occurred during the testing process. Though we do not have this information for the current dataset, across 240 toddler testing sessions, we have found that 97% of toddlers test within this margin. For adults we routinely obtain by calibration, and maintain during eye tracking, accuracy with error below 1 degree, and this is verified periodically as part of standard lab practices to assure data quality. Further details of the calibration and eye tracking procedures used were largely the same as those described in our prior paper's supplement [38].

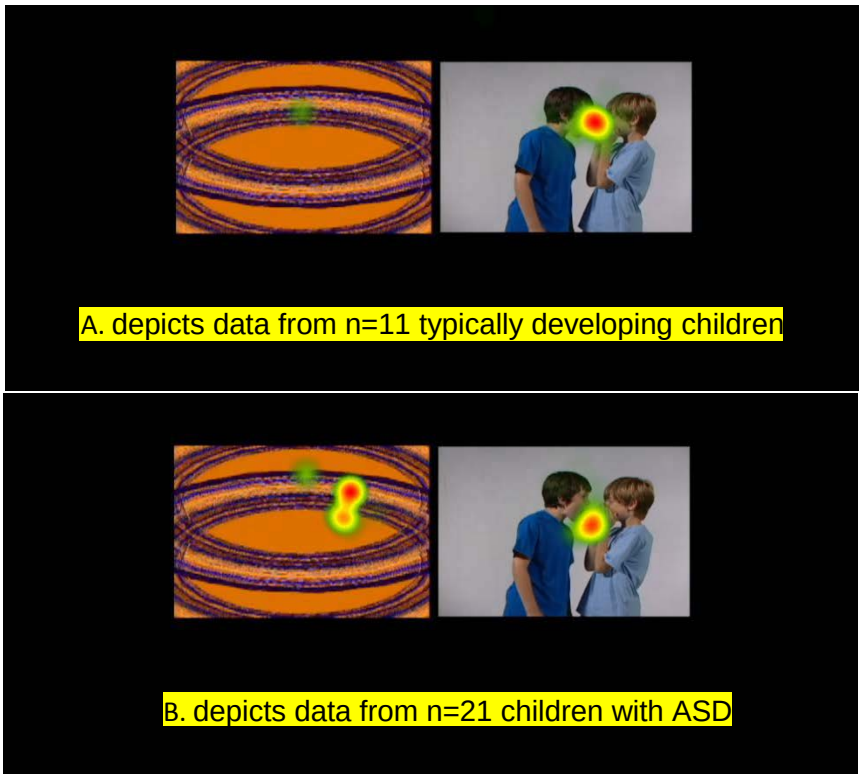

Figure 3.S2: Heat Maps of ASD and TD Gaze Fixation shows data "heat maps" drawn in Tobii Studio of convenience samples for illustrative purposes. The point of gaze fixation is usually near the AOI's center, so the outer perimeter of the AOI constitutes a large margin of error against spatial accuracy that falls outside the accuracy measurement the initial calibration would predict (e.g. due to drift).

The means and standard deviations for total looking time (in seconds, out of a possible total of 90 seconds) to either AOI by diagnostic group were as follows: ASD 65.6 (17.2); ASD Features 79.2 (11.5); DD 76.4 (15.1); TD 78.3 (13.0); Other 77.7 (15.9); TypSib 72.9 (21.6). There was a significant main effect ($F_{5,221}$ =5.8, p<.001) of diagnosis, and significant after Bonferroni correction post-hoc pairwise differences between ASD and the groups DD (p<.005), TD (p<.001) and Other (p<.05). There were no significant differences found between groups in total looking time to the original GeoPref test [38]. The difference found here can likely be attributed to the total duration of the Complex Social test video being 90 seconds, while the original GeoPref test's duration is 60 seconds.

Saccade Frequency to Geometric vs Complex Social Stimuli

Differences in saccades between clinical groups were reported for the original GeoPref test in our 2016 paper [38], therefore we report whether or not these effects were replicated with more complex social stimuli in the current study. For both the geometric and social stimuli, number of fixations per AOI was divided by sum of fixation time for that AOI to derive saccade frequency as saccades per second. Homogeneity of variance was confirmed then 1-way ANOVAs were performed (diagnostic group (6 levels) X saccades/sec (1 level)) for each AOI, and significant effects were followed by pairwise comparisons with Bonferroni correction. To confirm that differences in data quality were not impacting the reported results, ANCOVAs were performed as well, with 6 diagnostic groups as a fixed factor, saccade/sec as the dependent variable, and a data quality measure (percent of valid samples obtained) as a covariate. However, in this case, we found significant effects for the percent samples data quality measure. This implies that variable data quality between groups may be confounding true measurement of saccade rate, so results must be interpreted with caution.

Saccade Frequency of the 6 Groups to Geometric vs Complex Social Stimuli

For each toddler, each stimulus type (geometric and complex social) was considered separately to calculate saccades/sec. Saccade frequency data were subjected to additional scrutiny as this data quality can be impacted by data loss and lack of precision, while overall percent total fixation duration calculations are not as sensitive. Therefore, we excluded five subjects with saccades/sec values greater than two interquartile ranges from the upper quartile, indicating poor fixation filter performance [49]. While looking at geometric stimuli, there were no statistically significant differences in saccades/sec among the six diagnostic groups

($F_{5,216}$=.43, p=.8). However, ASD toddlers had significantly more saccades/sec when viewing social stimuli than did TD or DD toddlers ($F_{5,216}$=3.4, p=.005, partial eta-squared=.07; ASD vs DD, p<.005, Cohen's d=.63; ASD vs TD, p<.05, Cohen's d=.57). This is consistent with a faster rate of saccades occurring when the child is less interested in or attentive to the stimuli. But this could also reflect less ability to correctly measure saccades from some ASD children (also arguably due to being less interested or attentive).

Saccade Frequency Comparisons within ASD Subgroups

For geometric stimuli, the ASD GeoPref subtype had less frequent saccades, 1.33 saccades/sec, while the ASD SocPref subtype had more frequent saccades, 1.87 saccades/sec, ($t_{25.8}$=2.14, p<.05). For comparison, saccade rates to the same geometric stimuli reported in Pierce et al (2016) for these ASD subgroups were for ASD GeoPref 1.33 saccades/sec, and for ASD SocPref 1.94 saccades/sec. When viewing the new complex social stimuli, the opposite occurred: the ASD GeoPref group had more frequent saccades, 2.38 saccades/sec, and the ASD SocPref group has less frequent saccades, 1.61 saccades/sec, ($t_{18.4}$=-3.77, p=.001). See Figure S3. This is again consistent with faster rate of saccades when less attentive to the stimulus, but also potentially impacted by difficultly precisely measuring saccade frequency during less attentive behavior.

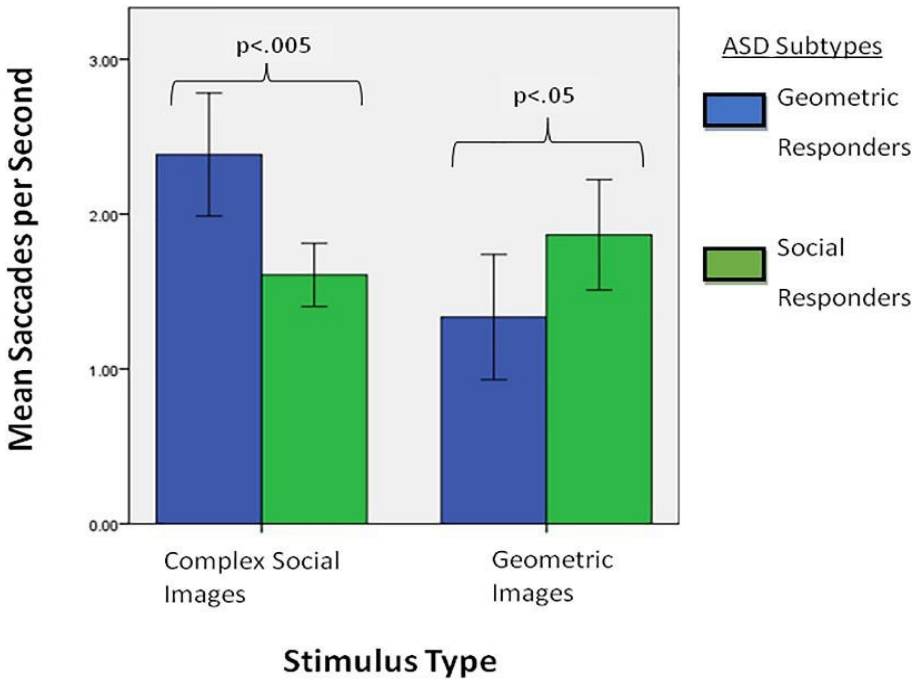

Figure 3.S3 Saccade Frequencies of ASD Subgroups to Complex Social & Geo Images
Bar graph illustrating comparison of saccade rate (saccades/second) between two ASD subgroups defined by their % Geo scores in the Complex Social GeoPref test: those who viewed geometric images more than 69% of the time, the Geometric Responder (GeoPref) subtype, and those who viewed social images more than 69% of the time, the Social Responder (SocPref) subtype. Group sizes were n=13 geometric responders and n=16 social responders. Error bars represent 95% confidence intervals.

Additional classification validation statistics

Table S1 shows classification validation statistics based on the cutoff for maximizing specificity with the Complex Social GeoPref test, 75% Geo looking time. At this cutoff, where specificity is 99% (more accurately, it is 99.78%), the PPV of the Complex Social GeoPref test slightly exceeds that of the original GeoPref test, at 92%. That is, on the original GeoPref test, at the 69% cutoff for %Geo the specificity is 99%, however the PPV is 90%. PPV is the likelihood that a given positive test is a true positive, and is therefore of particularly strong interest to clinicians.

Table 3.S1 Clinical Classification Performance, Complex Social GeoPref Test at 99% Specificity

%Geo threshold for positive test = 75%	ASD Only = True Positive N=227
True Positive	12
False Negative	64
False Positive	1
True Negative	150
Sensitivity	16%
Specificity	99%
Positive Predictive Value	92%
Negative Predictive Value	88%
Area Under ROC Curve	.74

Table S2 shows an additional potential usage for the Complex Social GeoPref test, ruling out an ASD diagnosis. That is, children with a %Geo score below 11% (i.e. a %Soc score above 89%) are very unlikely to have an ASD, with test classification performance shown below, where a lower score is more positive. This can be particularly useful as a second tier screen applied to toddlers who have already shown a few potentially concerning behaviors, and who may or may not have an urgent need for an autism focused evaluation. These values were derived by defining positive as falling into any diagnostic group other than ASD, and true positive by also have a %Geo score below 11%; negative refers to having an ASD diagnosis, and true negative to having an ASD and a %Geo score of 11% or greater. In this usage, if a test result is negative, it is inconclusive, because negative results are correct only 36% of the time (NPV=36% below). But if a test result is positive, it is valuable, because 92% of positive results are correct (PPV=92%), and, in this case, 22 toddlers (22 True Positives below) would be correctly identified as not at risk for ASD.

Table 3.S2 Clinical Classification Performance, Complex Social GeoPref Test for Ruling Out ASD

%Geo threshold for negative test = 10%	ASD Only = True Positive N=227
True Positive	22
False Negative	129
False Positive	2
True Negative	74
Sensitivity	15%
Specificity	97%
Positive Predictive Value	92%
Negative Predictive Value	36%
Area Under ROC Curve	.74

Table S3 shows validation classification statistics for detecting children with ASD without the inclusion of typically developing children (i.e. TD or TypSib groups), as distinguishing between ASD children and those with some sort of delay or other challenge that impacts behavior is typically the task facing clinicians. Also, our overall TD group contains some "control" participants, who would not be present in a sample from a non-research clinical setting, so this may be more reflective of a natural clinical sample.

Table 3.S3 Clinical Classification Performance, ASD vs DD and Other

69% Geo threshold	Complex Social GeoPref Test only	Complex Social and Original GeoPref Tests combined	Complex Social and Original GeoPref Tests combined
	ASD Only = True Positive vs DD & Other N=154	ASD Only = True Positive vs DD & Other N=77	ASD+ ASD Features = True Positive vs DD & Other N=87
True Positive	14	13	15
False Negative	62	24	32
False Positive	2	2	2
True Negative	76	38	38
Sensitivity	18%	35%	32%
Specificity	97%	95%	95%
Positive Predictive Value	88%	87%	88%
Negative Predictive Value	55%	61%	54%
Area Under ROC Curve	.75	.79	.79

Additional clinical severity in relation to eye tracking data information

Figure S4 (below) shows the significant correlation between %Geo scores on the Complex Social GeoPref test and ADOS scores for the entire group of 76 ASD study participants (Pearson's r=.46, p<.001).

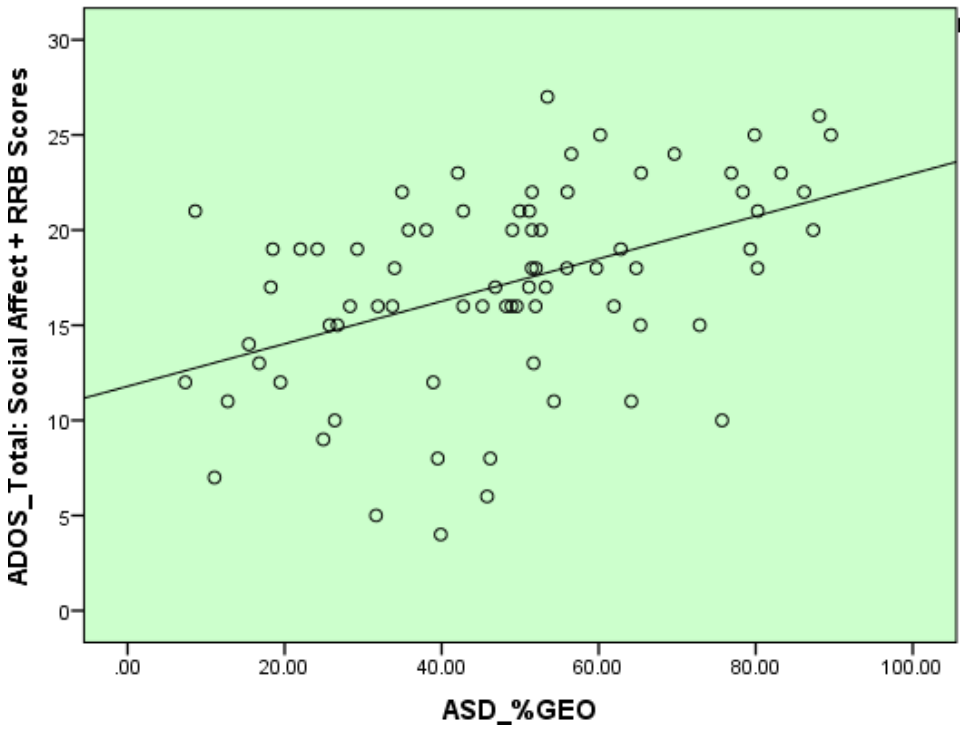

Figure 3.S4: Correlation between Complex Social Test %Geo and ADOS Scores in ASD. (n=76)

Experiment 2: Exploratory data on back to back usage of the original and Complex Social GeoPref tests

An additional 162 subjects participated in both GeoPref tests administered back to back on the same day. Between tests a break of several seconds was provided and a brief presentation of fixation crosses at known X-Y coordinates was used to confirm continued accuracy of gaze measurement.

Sixty-six of these participants (41%) were excluded from analysis due primarily to behavior incompatible with data collection (excessive movement, tantrums, etc) when exposed to the combined 2.5 minutes of GeoPref video presentation across two tests. In comparison, only 16% of participants from the main study, when only one GeoPref test was administered on a given day, were excluded. Of the remaining 96 subjects, 41 viewed the Complex Social test first and 55 viewed the original GeoPref test first. All were within the age range of our main study, 12 to 48 months. The diagnoses of the final 96 participants were 28 ASD, 4 ASD

features, 32 developmentally delayed, 22 typically developing, 5 typical siblings to ASD, 5 with other diagnoses.

Table S4 below shows classification results for the use of the two GeoPref tests back to back on the same day. While sensitivity, specificity and NPV remain fairly high, PPV drops below 2/3, meaning when a test results is positive there's more than a 1/3 chance that it is incorrect. This value, combined with the large percentage of subjects whose data cannot be used (41%), limits the clinical utility of administering the two GeoPref tests immediately back to back. In the future, the addition of musical sounds to increase willing attentiveness during back to back presentation of two GeoPref videos might be tested.

Table 3.S4 Clinical Classification Performance, Original and Complex Social GeoPref Tests Administered Immediately Back to Back

Positive = positive on either test **69% Geo threshold**	**ASD Only = True Positive** **vs all other groups** **N= 96**	**ASD+ ASD Features = True Positive** **vs all other groups** **N=96**	**ASD Only = True Positive** **vs DD & Other** **N=65**	**ASD+ ASD Features = True Positive** **vs DD & Other** **N=69**
True Positive	8	8	8	8
False Negative	20	24	20	24
False Positive	5	5	4	4
True Negative	63	59	33	33
Sensitivity	29%	25%	29%	25%
Specificity	93%	92%	89%	89%
Positive Predictive Value	62%	62%	67%	67%
Negative Predictive Value	76%	71%	62%	58%
Area Under ROC Curve	.76	.74	.74	.71

CHAPTER 4

Contents lists available at ScienceDirect

Behavioural Brain Research

journal homepage: www.elsevier.com/locate/bbr

Research report

EEG mu component responses to viewing emotional faces

Adrienne Moore[a,*], Irina Gorodnitsky[a], Jaime Pineda[a,b]

[a] University of California, San Diego, Department of Cognitive Science, 9500 Gilman Drive #0515, La Jolla, CA 92093-0515, USA
[b] University of California, San Diego. Department of Neurosciences, La Jolla, CA 92093-0662, USA

ARTICLE INFO

Article history:
Received 23 May 2011
Received in revised form 22 July 2011
Accepted 26 July 2011
Available online 3 August 2011

Keywords:
Mu rhythm
Facial expression
Simulation
SOBI
EEGLAB
Mirror neuron

ABSTRACT

Simulation theories for the perceptual processing of emotional faces assert that observers recruit the neural circuitry involved in creating their own emotional facial expressions in order to recognize the emotions and infer the feelings of others. The EEG mu rhythm is a sensorimotor oscillation hypothesized to index simulation of some actions during perceptual processing of these actions. The purpose of this research was to extend the study of mu rhythm simulation responses during perceptual tasks to the domain of emotional face perception. Subjects viewed happy and disgusted face photos with empathy and non-empathy task instructions while EEG responses were measured. EEG components were isolated and analyzed using a blind source separation (BSS) method. Mu components were found to respond to the perception of happy and disgusted faces during both empathy and non-empathy tasks with an event-related desynchronization (ERD), activation that is consistent with face simulation. Significant differences were found between responses to happy and to disgusted faces across the right hemisphere mu components beginning about 500 ms after stimulus presentation. These findings support a simulation account of perceptual face processing based on a sensorimotor mirroring mechanism, and are the first report of distinct EEG mu responses to observation of positively and negatively valenced emotional faces.

1. Introduction

Social interaction relies heavily on nonverbal communication, especially through facial expressions. We make critical inferences about the feelings, motivations, and intentions of others based on observing facial emotion. Simulation theories for the processing of emotional faces assert that observers activate the sensorimotor representations involved in creating their own emotional facial expressions, in order to recognize the emotions and infer the feelings of others [1,2].

Several different forms of facial expression simulation have been identified. One of these is facial mimicry, activation of the facial muscles used to produce a given emotion expression in response to seeing the expression on another face [3,4]. Mimicry has been shown to influence how we feel, for example, to increase liking between interaction partners, and thus to facilitate successful social interactions [5]. A second type of facial simulation is facial mirroring, activation of neural substrates for expression production that is not for the sake of mimicry but rather for producing an "offline" simulation of what is observed [6,7]. Facial mirroring, which refers primarily to the mirroring of the action of generating an emotional facial expression, is believed to be important for making inferences about how others are feeling [2]. This involves shared premotor representations utilized both for the action of generating or imitating an emotional facial expression oneself and for the perceptual processing of another's face [6,7]. Mimicry and mirroring systems are believed to be distinct, because while viewing the actions of others one's own primary motor cortex (M1) is not usually active [8], and functional magnetic resonance imaging (fMRI) studies of facial mirroring responses have shown premotor but not primary motor cortex activity [6,7]. Some researchers assert that having two distinct systems for mirroring and mimicry may help us distinguish our own actions and sensations from those of others whom we simulate [9].

Another aspect of mirroring in response to faces is the simulation of the emotional feeling conveyed by the face, in addition to the simulation of the facial configuration. Emotion mirroring involves recruiting the cortical activation involved in one's own experiences of an emotion or bodily sensation during the perceptual processing of the same feelings expressed in others [9], and co-occurs with facial mirroring [6]. There is evidence for neural mirroring of an observed feeling state in response to facial expressions across a number of specific feeling states. These include shared neural substrates for feeling and for observing disgust [10], and for feeling and observing pain [11]. Together these simulation processes provide a plausible foundation for the human capacity for inferring and responding to the feelings and experiences of others [12].

* Corresponding author. Tel.: +1 858 373 8705; fax: +1 858 534 1128.
E-mail addresses: armoore@ucsd.edu, adriennermoore@yahoo.com (A. Moore).

0166-4328/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbr.2011.07.048

Simulation of observed actions can be investigated by measuring the electroencephalogram (EEG) mu rhythm during action observation [13]. The mu rhythm is a cortical oscillation generated by somatotopically organized sensorimotor cortex near the central sulcus, with distinct oscillations occurring for processing related to different body parts [14]. The neural generators of the mu signal are said to "idle" synchronously when off task, which is typically measured as high 8–13 and 15–25 Hz EEG power at central electrode sites, at rest with eyes open [15]. The mu rhythm is activated, or desynchronized, resulting in suppression of mu spectral power, by movement and movement preparation [16]. The mu rhythm is similarly desynchronized and suppressed by the observation or the imagination of movements of body parts (hands and feet have been primarily studied), linking the mu rhythm to both action and action simulation [13].

Simulation theories of person perception and social cognition were fostered by the study of mirror neurons in monkeys and then of the human mirror neuron system. Individual macaque neurons in premotor region F5 fire to both object-directed hand actions, to ingestive and communicative mouth and lip actions, and to their observation [17]. The human inferior frontal gyrus (IFG), which is believed to be the homolog to macaque F5, has been shown to be similarly activated by perception of actions [18]. This activation appears to facilitate action understanding, inferring goals and intentions when observing people's behavior. For example, premotor mirror neuron regions, active during both execution and observation of an action, are more strongly activated by perception of grasping hands within a meaningful context (drinking or cleaning up) when compared to perception of grasping hands without a contextual scene or to perceiving the context alone [19].

A number of studies infer from functional similarities between the mu rhythm and the human mirror neuron system (MNS) that changes in EEG mu (at least the 8–13 Hz component, where this research has focused) reflect "downstream" modulation of sensorimotor cortex neurons by premotor cortex mirror neurons, potentially located in BA 44 [13,20]. Consistent with this, using repetitive transcranial magnetic stimulation (rTMS) to disrupt activity in the left inferior frontal gyrus affects the modulation of mu rhythms over sensorimotor cortex [21]. A canonical mirror neuron triggering stimulus, the observation of an object-directed hand grasping motion, has been shown to decrease mu power relative to simple hand extension and to objectless grasping gestures [22,23]. This shows that the mu rhythm shares the mirror neuron system's preferential activation for goal-directed actions [17]. Further linking mu rhythm simulation to the mirror neuron system, it has been shown that observation of point-light biological motion of full-body gestures (e.g. jumping jacks) both activates premotor MNS areas according to fMRI research [24], and causes EEG mu power suppression [25]. Linking mu power to critical social processes, stimuli varying in degree of sociality (movies of social ball tossing games) have been asserted to suppress the mu rhythm accordingly [26]. Finally, mu suppression has also been correlated with accuracy on social-perceptual tasks, which involve inferring mental states from bodily expressions, but not on social-cognitive tasks, which are linked to language and theory building.

Regarding mu and simulation of faces in particular, very little research has been done. Mental imagery of orofacial movements (tongue movement and lip movement) has been shown to affect the EEG mu rhythm in a manner suggesting simulation in the form of recruitment of motor cortical areas for movement that is imaginary, not actual [27,28]. But mu power changes in response to emotional face observation have not been previously characterized.

This current study predicted that because viewing emotional faces involves simulation of the emotional facial expression, it also elicits a mu rhythm desynchronization. To test this, we compared mu responses to positively and negatively valenced faces, specifically, to happy and disgusted faces. Because simulation mechanisms are known to differ for different emotion expressions [9,29], we predicted that the mu desynchronization responses to happy and to disgusted faces would be distinct. Because evidence shows that the right hemisphere is preferentially involved in perceptual processing of emotion and emotional faces, we predicted these differences may be right lateralized [30,31]. Finally, we also compared the influence of task conditions that do and do not instruct subjects to attempt to empathize with the perceived emotions. This allowed us to investigate whether the mu face simulation is automatic, or whether deliberate attention to the emotion expressed and an attempt to put oneself "into the shoes" of the observed person influence the mu response.

2. Materials and methods

2.1. Subjects

Thirty undergraduate students were recruited through UCSD courses and compensated for participation with course extra credit. Subjects' vision was normal or corrected to normal. Exclusion criteria included history of neurological disease and current use of psychotropic medications or stimulants other than caffeine. Subjects were also excluded if they scored above 17 (borderline clinical depression) on the beck depression inventory, or if lack of fluency in English interfered with comprehending the instructions. Five subjects were not included in the final study, due to data recording problems. Of the remaining twenty-five subjects, twenty-two (11 females and 11 males) yielded clear mu components as defined in Section 2.4.3. The human study protocol was approved by the Institutional Review Board at UCSD and therefore has been performed in accordance with the ethical standards of the 1964 Declaration of Helsinki. All subjects gave their informed consent prior to the beginning of the experiment.

2.2. Stimuli and experimental design

Subjects viewed six blocks of 40 photos: four blocks of face photos (happy and disgusted, with and without the explicit instruction to empathize), one block of photos of buildings, and one block of static visual noise images. The order in which subjects completed the conditions was pseudo-randomized with all orders approximately equally represented in the final set of subjects, such that the 22 final subjects represented 18 different orders. Finally, at the end of the experiment subjects completed the balanced emotional empathy scale (BEES), a self-report tool for indexing individual differences in dispositional emotional empathy [32].

Face images were greyscaled photos from a validated facial affect stimulus set (the MacArthur Foundation Research Network EEBD NimStim set), which was created by actors coached by a FACS (facial affect coding system) expert [33]. Each block of faces contained 40 unique photos of one emotion (happy or disgusted). The photos depicted both genders and three ethnicities (European–American, African–American, and Asian–American) in random order. Face photos from particular actors and actresses were counterbalanced to appear paired with empathy and non-empathy task instructions with equal frequency.

In the empathy conditions, subjects were instructed to try to experience the emotions felt and expressed by the photographed people, and then to rate how successful they believe they were at empathizing with each on a 1–5 Likert scale. To ascertain whether subjects experienced a change in emotion congruent with the observed faces during the empathy tasks, subjects reported their mood after each block of 40 empathy trials using the PANAS-X (positive and negative affect scale-expanded). The PANAS-X is a well-validated, self-report measure of transient change in mood, consisting of adjectives and short phrases describing potential feelings and emotions in response to which subjects reported to what extent these terms described how they felt at that moment [34]. In the non-empathy conditions, subjects were asked to rate how attractive they found each face on a 1–5 Likert scale, a task, which did not require directly attending to the expressed emotion. In the buildings condition, subjects were asked to rate how well they liked each building on a 1–5 Likert scale. The constraints on order randomization were that two conditions of the same task (empathy or non-empathy) were always adjacent to one another, and the final block of photos viewed was never an empathy condition, to avoid administering the final PANAS-X self-report at the end the experiment when mood is perturbed.

2.3. EEG data acquisition

Comfortably seated in a soundproof and electrically shielded recording booth, subjects were told to fixate gaze at the center of the monitor as much as possible, to minimize movement of facial muscles, to blink only when rating with a button press, and to wait until prompted by a question appearing on the monitor to respond with the button press (see Fig. 1). Care was taken to instruct subjects not to move more than necessary, stressing the potential for motion artifact.

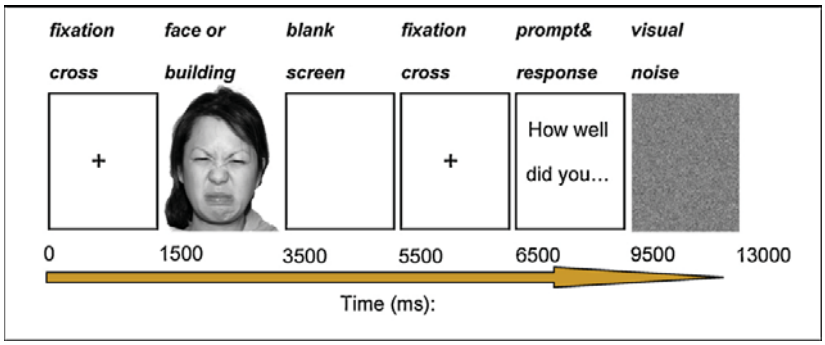

Fig. 1. Timing of individual trials. Stimuli were presented for 2000 ms.

Continuous EEG was recorded using a NeuroScan system (NeuroScan, Inc., Herndon, VA) from 13 electrodes placed according to the International 10–20 System (O1, O2, T5, T6, P3, PZ, P4, C3, CZ, C4, F3, FZ, and F4) referenced to electronically linked mastoid electrodes. Additionally, bipolar electrooculogram (EOG) was recorded from four electrodes to monitor blinks and eye movements (positioned vertically at the supraorbital ridge and lower outer canthus of the left eye, and horizontally at the middle outer canthi of left and right eyes). Impedances were set below 10 kΩ (usually below 5 kΩ) for HEOG and VEOG, and below 5 kΩ for mastoids and cap electrodes. Data were sampled at 500 Hz and filtered to the .05–30 Hz band by the NeuroScan acquisition software.

2.4. EEG data analysis

The EEGLAB Matlab Toolbox [35] was used as the platform for data analysis, with some integrated customized routines, including eSOBI, the second-order blind identification (SOBI) algorithm for epoched data [http://sites.google.com/site/bioanalyze/]. We used blind-source separation (BSS) to reduce noise and artifacts in the EEG data and to extract the mu rhythm components.

2.4.1. Preprocessing

Data records from each subject were highpass filtered above 4 Hz, yielding 4–30 Hz bandpassed data. Data epochs time-locked to the presentation of a single photo were extracted from 1500 ms before stimulus presentation to 2000 ms after stimulus presentation. Mean baseline values were removed from each epoch. Trials containing button presses incorrectly made during the epoched time period were deleted (resulting in 191–200 total epochs per subject). Data epochs from the visual noise condition were removed and the remaining epochs for each subject were concatenated. The reason visual noise was excluded was that it elicited strong, occipital

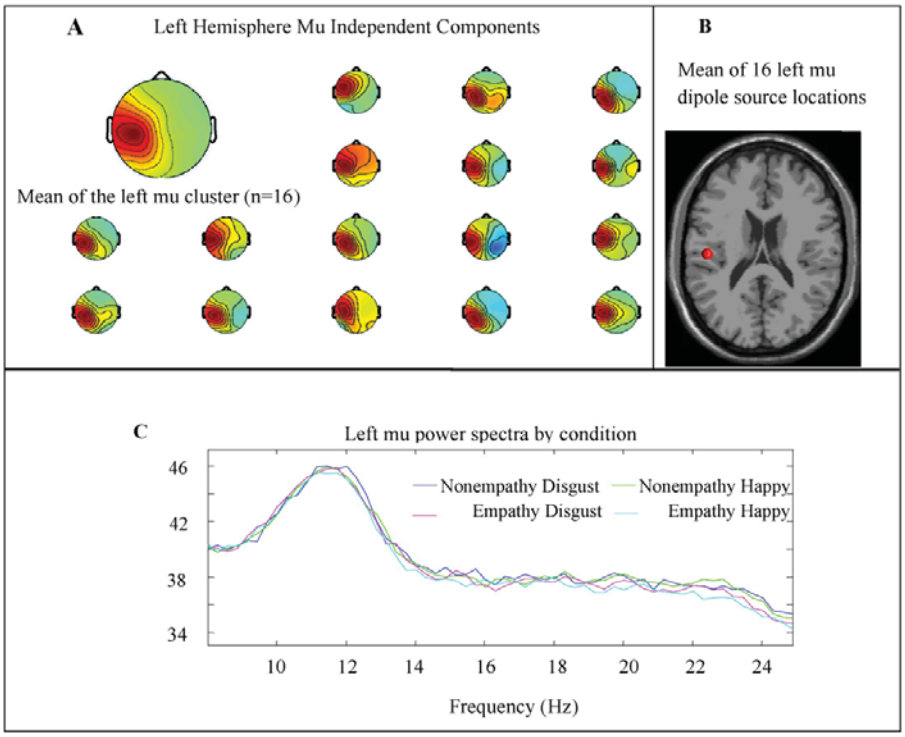

Fig. 2. Left mu cluster: (A) Mean and individual topographic scalp maps, (B) Cluster mean DIPFIT dipole location, (C) Cluster mean power spectrum, $10 \times \log_{10}(\mu V^2/Hz)$), for each condition.

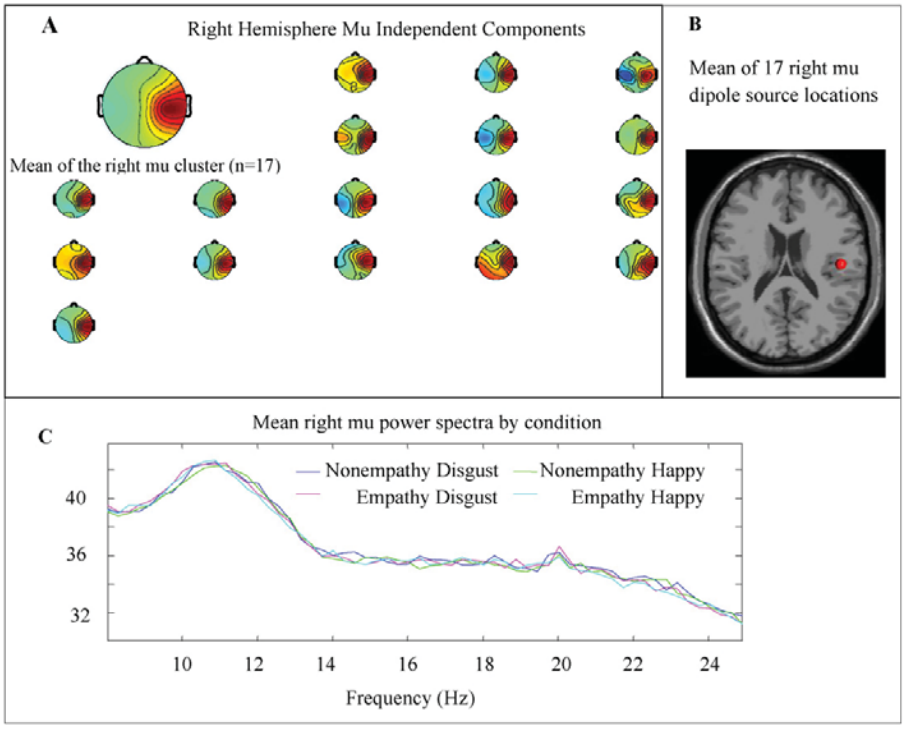

Fig. 3. Right mu cluster: (A) Mean and individual topographic scalp maps, (B) Cluster mean DIPFIT dipole location, (C) Cluster mean power spectrum, $10 \times \log_{10}$ (μV^2/Hz)), for each condition.

8–13 Hz alpha activity apparent in the raw EEG record, which can occasionally alter the robustness of source separation of other, weaker cerebral sources.

2.4.2. Component separation

SOBI, a well-validated BSS method, was used to separate the preprocessed EEG signals into independent components. BSS attempts to separate linearly mixed signals without information about the mixing process. The SOBI algorithm, introduced by Belouchrani et al. [36], attempts to perform the joint approximate diagonalization of a set of time-lagged covariance matrices obtained from the data. SOBI has been found to be effective in removing EOG and EMG artifacts, noise, and also for separating EEG cerebral sources, including mu components [37–40]. SOBI is also computationally fast and can work well on short data records.

The eSOBI algorithm is a modification of SOBI for processing epoched data.

The eSOBI algorithm was applied three times as follows. In the first step, eSOBI was used similarly to the procedures described by Ng and Raveendran [38] to extract and remove components that contained a large portion of noise and artifacts, including ocular artifacts. Next, "cleaned" EEG data were reconstructed from the remaining non-artifactual components. In the second step, eSOBI was applied to the cleaned EEG data to identify trials in which subject motion distorted EEG components. After rejecting those trials, cleaned EEG data were once again reconstructed and eSOBI was applied a third time. The mu rhythm components were identified from the third decomposition. Sequential application of eSOBI yielded data with improved signal to noise ratio, by unmixing and removing standard EEG artifacts from the brain activity data. This overall approach was found to yield the most robust results for our data.

2.4.3. Identifying mu components

To identify mu rhythm components, we first used individual topographic scalp maps (Figs. 2A and 3A) to identify all centrally located generators, with dipolar structure, and with the focus lateralized left or right of the midline. However, topographic projections depend on the orientation of the electric field, so sources from other parts of the cortex may occasionally appear as emanating from the sensorimotor region in a topographic map. One solution is to localize the cortical generators of selected components to verify their foci. Therefore, as our second validation step we performed such localization using the equivalent dipole DIPFIT function and standard Boundary Element Model (MNI) head model, with the warp montage function to co-register the electrode locations with the head model, followed by coarse and fine fitting [35] (Figs. 2B and 3B). The localization identified one left lateralized component, which was located outside the central cortical region, and it was excluded from further analysis. In the third validation step, spectral analysis was performed on the remaining components. All of the components that passed the second validation step had peak power in the alpha (8–13 Hz) range (Figs. 2C and 3C), as is characteristic of mu. This analysis identified 22 subjects with at least one clear mu component. Of those 22, 11 had both left and right hemisphere mu components. No subjects had more than one mu component per hemisphere, indicating that the identified components captured the entire left or right mu rhythm responses extracted for each subject. The final mu clusters contained 17 right lateralized and 16 left lateralized mu components (Figs. 2A and 3A).

2.4.4. Spectral and statistical analyses

Event related spectral perturbations (ERSPs), deviations in spectral power relative to a baseline, were calculated for each component in the left and right mu clusters using built-in EEGLAB procedures [35] as follows. A time-frequency decomposition was computed for each individual condition using wavelets with Morlet tapers, and the deviations in log spectral power in each time-frequency bin were then computed, relative to the mean of the log spectral power of the 1500 ms pre-stimulus baseline. To compare responses for specific experimental conditions, the common baseline was calculated across those test conditions using EEGLAB, and the component ERSP values were adjusted for the common baseline for each test.

To assess statistical differences, nonparametric resampling methods available in EEGLAB were used [41]. A bootstrap resampling methods was used to test whether ERSP deviations in spectral power in the post-stimulus interval were significantly larger relative to the pre-stimulus period for each subject and each separate condition. The statistical differences across the four face conditions were analyzed for the right and the left clusters by comparing ERSP values in each time-frequency bin for that cluster using repeated measures permutation comparison of the 2X2 (empathy/non-empathy X happy/disgusted) design. To address the increased probability of false discoveries in multiple hypotheses testing, all ERSP results were corrected using Benjamini and Hochberg [42] false discovery rate correction with an alpha value of .05.

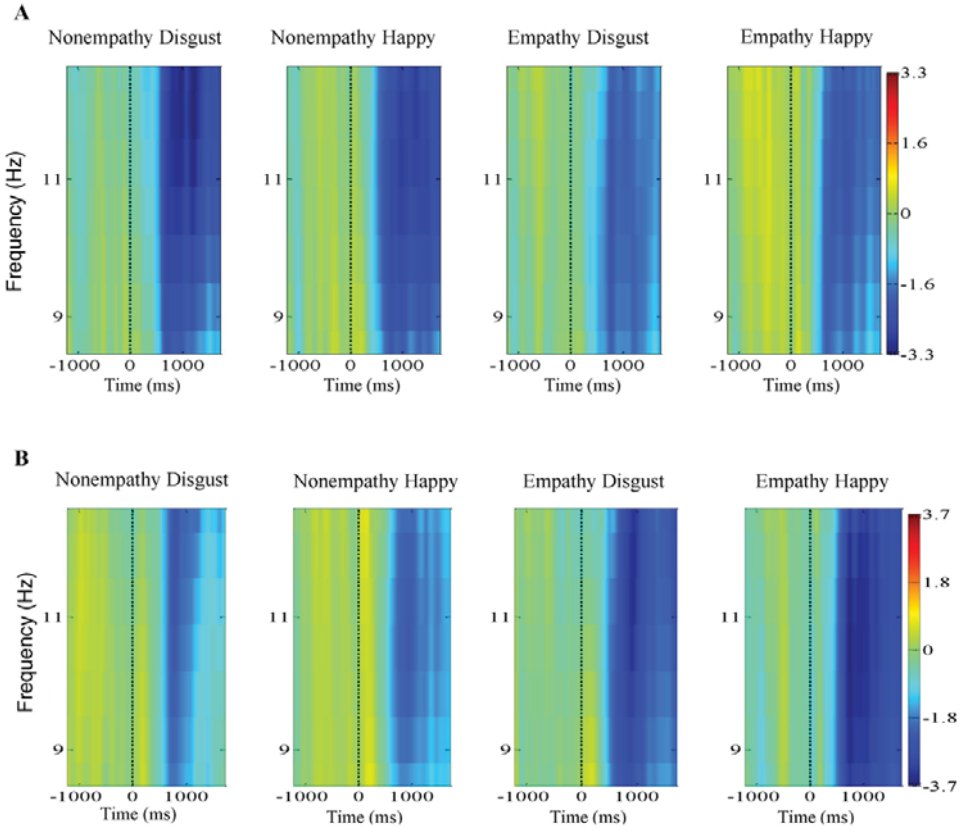

Fig. 4. (A) Left mu cluster, event related spectral perturbations (8–13 Hz power, dB), (B) Right mu cluster, event related spectral perturbations (8–13 Hz power, dB).

3. Results and discussion

3.1. Emotion and empathy self-report results

Mood measurement (the PANAS-X) was administered two times to each subject, following the two empathy tasks. The differences (first score minus second score) between mood change for subjects who would be predicted to have a mood score drop if empathizing (empathized with happy first, then with disgust) and subjects who would be predicted to have a mood score increase if empathizing (empathized with disgust first and then with happy) were compared. 90% of the subjects who empathized with happy faces first and then disgusted faces experienced a drop in mood, with a mean drop of 8.6 points. 62.5% of the subjects who empathized with disgusted faces first and then happy faces experienced an improvement in mood, with a mean increase of 2.75 points. This difference was statistically significant (2 tailed, unequal variance t-test, $t(15.7) = -3.59$, $p < .003$, after confirming data does not differ from Gaussian, Kolmogorov-Smirnov $z = 355$, $p = \sim 1.0$), confirming that the empathy task successfully modulated subject mood. Though mood responses were not measured following the non-empathy conditions, it is possible that subjects also experienced an empathetic mood convergence during the non-empathy conditions due to observing emotional faces.

Individual differences in trait emotional empathy reported through the BEES questionnaire were calculated as well, with the prediction that the trait empathy measure may be correlated with how much self-reported mood was perturbed in the empathy conditions and/or with the magnitude of the mu face response ERSPs. However, no significant correlations were found between individual differences in trait empathy and amount of mood change due to observing emotional faces or magnitude of mu EEG response.

3.2. EEG component separation results

To cluster mu components together we used mu power spectra, topographic maps, and equivalent dipole source localizations as criteria for inclusion of a component in the mu clusters. Mu components were found in the left and/or the right hemispheres of most subjects (22 of 25). In Figs. 2A and 3A it can be observed that the centers of the topographic projections of components vary slightly more for the left than the right cluster. The underlying cause may be poorer performance for source separation of left mu responses due to weaker responses in the left than the right hemisphere during emotional face perception [30].

3.3. EEG spectral and statistical results

Event-related perturbations in spectral power time-locked to stimulus presentation for the left and right mu component clusters were used to compare experimental conditions as described in Section 2.4.4. ERSPs are referred to as event-related desynchronizations (ERDs) if the direction of post-stimulus change in power is decreasing, and as event-related synchronizations (ERSs) if the direction of post-stimulus change in power is increasing. The ERD corresponds with mu power suppression, which has been reported

in response to movement and movement simulation, while the ERS corresponds to mu power enhancement. All significant mean ERSPs from all subjects and all face viewing conditions in both clusters except for one mean response were ERDs, not ERSs, consistent with the face simulation hypothesis.

3.3.1. Buildings to faces ERSP comparison

To compare mu ERSP responses to viewing emotional faces with mu ERSP responses to the control viewing buildings, for each subject we calculated the mean response for each condition across 40 trials. Observing faces elicited more statistically significant ERD responses than observing buildings in both the left and right mu clusters, consistent with a face simulation mechanism indexed by mu suppression. In the left mu cluster, buildings elicited a statistically significant ($p < .05$) ERD response in 50% of subjects, while faces elicited a significant ERD response in 81.3%. On the right, buildings elicited a significant ERD response in 47.1% of subjects, while face elicited a significant ERD in 76.5%. We confirmed that the difference between percentage of subjects who responded to faces and percentage of subjects who responded to buildings in each cluster was significant with a Wilcoxon signed ranks test ($z = -2.236$, $p = .025$, both clusters). While this confirmed significant mu suppression to viewing emotional faces relative to the control condition, further studies are needed to understand the nature of the mu response to observing buildings. Future research should determine whether the mu response to buildings indicates a baseline responsivity of sensorimotor EEG components to any complex visual stimuli observed for the sake of a rating task. Following this confirmation that faces elicited significantly greater mu suppression than the control buildings, we then compared in detail differences between the responses to the four face observation conditions.

3.3.2. Face conditions ERSP comparison

The mean ERSP values computed as described in Section 2.4.4 above are displayed in Fig. 4. ERDs were observed beginning approximately 500 ms after stimulus presentation in both the left and the right mu clusters in the 8–13 Hz range in response to observing both happy and disgusted faces, with and without the explicit instruction to empathize with the faces. As ERD responses indicate activation of the mu rhythm and are characteristic of real and imagined bodily movement, this suggests simulation of the action of producing a facial expression in response to observing both positively and negatively valenced faces.

Though faces elicited essentially the same number of significant ERDs from the left and right hemispheres (from 81% of subjects on the left and from 77% of subjects on the right), there were significant differences between the four face processing condition ERDs only on the right. The results from non-parametric, permutation-based statistical comparison of the right hemisphere ERSP responses to the four face processing conditions are shown in Fig. 5. B. A main effect of facial emotion observed (happy vs. disgusted) was found in the right mu clusters ($p < .05$). This lateralization of effects is consistent with many reports of different roles for the left and right hemispheres in processing emotional faces, and with many reports of right hemisphere dominance for face processing and emotion processing [1,30,31]. The time course of the post-stimulus response main effect of facial emotion observed involves a significantly greater ERD to disgusted faces than to happy faces at around 500 ms post-stimulus presentation, followed by a larger, significantly greater ERD to happy faces than to disgusted faces at around 600 ms, 1000 ms and 1500 ms (see Fig. 5A and B). The time course of these differences indicates more rapid face simulation for negatively valenced, disgust faces, but an overall more extensive simulation response for positively valenced, happy faces.

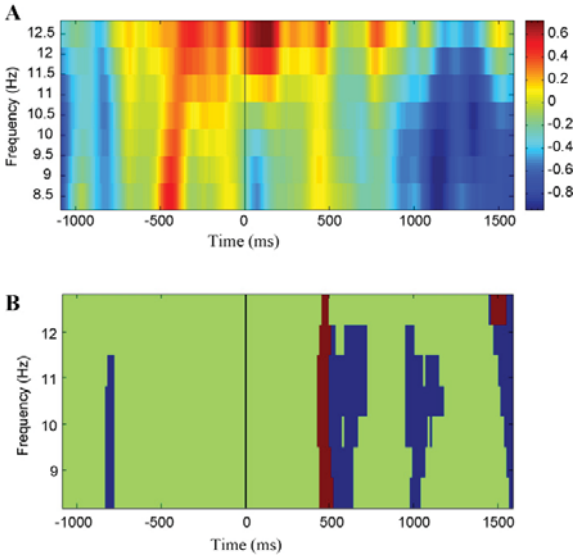

Fig. 5. (A) Main effect of emotional facial expression, right hemisphere mu ERSP power. Response to happy faces minus response to disgusted faces, right mu cluster power, in dB (8–13 Hz). Shows the mean ERSP power collapsed across the empathy and non-empathy conditions ((EH-ED + NH-ND)/2). Redder regions indicate ERSP power to happy > ERSP power to disgust, therefore, post-stimulus ERD to disgust > ERD to happy. Bluer regions indicate that ERSP to happy < disgust, therefore, ERD to happy > ERD to disgust. (B) Significant differences by emotional facial expression, right hemisphere mu ERSP. Shows the regions where ERSP differences between happy and disgust are statistically significant. In red regions the ERD to disgusted faces is significantly greater than the ERD to happy. In blue regions the ERD to happy faces is significantly greater than the ERD to disgusted faces. $p < .05$ after FDR correction at all red and blue data points.

At around −800 ms pre-stimulus, the right hemisphere power briefly differs significantly between happy and disgusted face conditions, with higher power in the happy face conditions ($p < .05$). Because the stimuli were presented in blocks to facilitate mood convergence in one direction across blocked trials, subjects were able to anticipate the upcoming facial expression. These pre-stimulus differences by face type are attributed to anticipatory simulation of facial expressions, as mental imagery is another class of simulation known to be reflected in mu power changes [27,28].

After FDR correction for multiple comparisons across time and frequency bins, no significant differences were found in the mu component responses when comparing empathy and non-empathy conditions ($p > .05$). Absence of differences between the mu responses in empathy and non-empathy conditions in this study may indicate that the mu ERD reflects an automatic simulation process which is not influenced by the deliberate attempt to empathize. Alternately, it is possible that this analysis was not sensitive enough to detect significant differences by task, in part due to the correction for multiple comparisons across a large number of time and frequency bins (1400), which increases the probability of false negatives. Future research could address these alternatives.

3.4. Mirroring and mimicry

Viewing emotional faces has been shown to elicit a facial mimicry response measured by EMG recording from electrodes placed above the facial muscles responsible for generating a particular expression [3] [43]. This mimicry occurs both within 500–1000 ms [44], and also at slower time scales [45,46]. Facial

mimicry responses could contribute to the mu suppression effects reported here.

However, evidence also exists indicating that mu power changes while observing face stimuli do not necessarily reflect this covert muscle movement. An MEG study of observation of faces producing speech and non-speech lip forms while EMG was collected reported activity in BA 44 and in sensorimotor areas, indicating simulation, but found no significant facial EMG activity [47]. This is consistent with other studies of mu simulation responses and EMG. For example, in a study wherein subjects imagined hand movements in order to move a cursor by means of a brain–computer interface, hand EMG activity was reported to be very low. Further, the correlation between cursor target position and EMG was reported to be much lower than that between target position and mu EEG [48]. This supports the idea that mu reflects a simulation process that is distinct from motor processes that produce mimicry's movement. Future research on mu suppression in response to emotional face perception should attempt to distinguish between mu ERD effects due to facial mimicry and mu ERD effects due only to neural mirroring. Simultaneous measurement of facial EMG and EEG responses while subjects observe emotional faces may accomplish this.

4. Conclusion

This study extends EEG mu mirroring research to the domain of emotional face perception by identifying the mu responses to viewing happy and disgusted facial expressions. Mu component event-related desynchronization responses occurred to observation of both disgusted and happy faces with and without the deliberate attempt to empathize with the emotion viewed. This event-related decrease in mu component power is believed to indicate simulation of the action of producing an observed facial expression, consistent with previous accounts of mu power suppression responses to both action execution and action perception in other domains [23,25]. In the right hemisphere, the mu component responses to observation of these positively and negatively valenced emotional faces were distinct. Specifically, at around 500 ms disgusted faces elicited a more robust ERD, with a stronger ERD to happy faces subsequently. The findings reported here support a simulation theory of face processing based on a sensorimotor mirroring mechanism, wherein first-person action execution and third-person action perception share common neural substrates.

Acknowledgements

Special thanks to Dr. Andrea Chiba and the Temporal Dynamics of Learning Center for mentoring and support.

Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the stimulus set.

References

[1] Adolphs R. Recognizing emotion from facial expressions: psychological and neurological mechanisms. Behav Cogn Neurosci Rev 2002;1:21–62.
[2] Adolphs R. How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. Brain Res 2006;1079:25–35.
[3] Dimberg U. Facial reactions to facial expressions. Psychophysiology 1982;19:643–7.
[4] Hess U, Blairy S. Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. Int J Psychophysiol 2001;40:129–41.
[5] Chartrand TL, Bargh JA. The chameleon effect: the perception-behavior link and social interaction. J Pers Soc Psychol 1999;76:893–910.
[6] Carr L, Iacoboni M, Dubeau MC, Mazziotta JC, Lenzi GL. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. Proc Natl Acad Sci USA 2003;100:5497–502.
[7] Leslie KR, Johnson-Frey SH, Grafton ST. Functional imaging of face and hand imitation: towards a motor theory of empathy. Neuroimage 2004;21:601–7.
[8] Gazzola V, Keysers C. The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. Cereb Cortex 2009;19:1239–55.
[9] Bastiaansen JA, Thioux M, Keysers C. Evidence for mirror systems in emotions. Philos Trans R Soc Lond B: Biol Sci 2009;364:2391–404.
[10] Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G. Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. Neuron 2003;40:655–64.
[11] Saarela MV, Hlushchuk Y, Williams AC, Schurmann M, Kalso E, Hari R. The compassionate brain: humans detect intensity of pain from another's face. Cereb Cortex 2007;17:230–7.
[12] Gallese V, Keysers C, Rizzolatti G. A unifying view of the basis of social cognition. Trends Cogn Sci 2004;8:396–403.
[13] Pineda JA. The functional significance of mu rhythms: translating seeing and hearing into doing. Brain Res Brain Res Rev 2005;50:57–68.
[14] Pfurtscheller G, Neuper C, Andrew C, Edlinger G. Foot and hand area mu rhythms. Int J Psychophysiol 1997;26:121–35.
[15] Pfurtscheller G, Stancak Jr A, Neuper C. Event-related synchronization (ERS) in the alpha band—an electrophysiological correlate of cortical idling: a review. Int J Psychophysiol 1996;24:39–46.
[16] Neuper C, Wortz M, Pfurtscheller G. ERD/ERS patterns reflecting sensorimotor activation and deactivation. Prog Brain Res 2006;159:211–22.
[17] Rizzolatti G, Craighero L. The mirror-neuron system. Annu Rev Neurosci 2004;27:169–92.
[18] Kilner JM, Neal A, Weiskopf N, Friston KJ, Frith CD. Evidence of mirror neurons in human inferior frontal gyrus. J Neurosci 2009;29:10153–9.
[19] Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G. Grasping the intentions of others with one's own mirror neuron system. PLoS Biol 2005;3:e79.
[20] Oberman LM, Hubbard EM, McCleery JP, Altschuler EL, Ramachandran VS, Pineda JA. EEG evidence for mirror neuron dysfunction in autism spectrum disorders. Brain Res Cogn Brain Res 2005;24:190–8.
[21] Keuken MC, Hardie A, Dorn BT, Dev S, Paulus MP, Jonas KJ, et al. The role of the left inferior frontal gyrus in social perception: an rTMS study. Brain Res 2011;1383:196–205.
[22] Muthukumaraswamy SD, Johnson BW. Changes in rolandic mu rhythm during observation of a precision grip. Psychophysiology 2004;41:152–6.
[23] Muthukumaraswamy SD, Johnson BW, McNair NA. Mu rhythm modulation during observation of an object-directed grasp. Brain Res Cogn Brain Res 2004;19:195–201.
[24] Saygin AP, Wilson SM, Hagler Jr DJ, Bates E, Sereno MI. Point-light biological motion perception activates human premotor cortex. J Neurosci 2004;24:6181–8.
[25] Ulloa ER, Pineda JA. Recognition of point-light biological motion: mu rhythms and mirror neuron activity. Behav Brain Res 2007;183:188–94.
[26] Oberman LM, Pineda JA, Ramachandran VS. The human mirror neuron system: a link between action observation and social skills. Soc Cogn Affect Neurosci 2007;2:62–6.
[27] Spiegler A, Graimann B, Pfurtscheller G. Phase coupling between different motor areas during tongue-movement imagery. Neurosci Lett 2004;369:50–4.
[28] Pfurtscheller G, Brunner C, Schlogl A, Lopes da Silva FH. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. Neuroimage 2006;31:153–9.
[29] Lee TW, Josephs O, Dolan RJ, Critchley HD. Imitating expressions: emotion-specific neural substrates in facial mimicry. Soc Cogn Affect Neurosci 2006;1:122–35.
[30] Borod JC, Cicero BA, Obler LK, Welkowitz J, Erhan HM, Santschi C, et al. Right hemisphere emotional perception: evidence across multiple channels. Neuropsychology 1998;12:446–58.
[31] Killgore WD, Yurgelun-Todd DA. The right-hemisphere and valence hypotheses: could they both be right (and sometimes left)? Soc Cogn Affect Neurosci 2007;2:240–50.
[32] Mehrabian A, Epstein N. A measure of emotional empathy. J Pers 1972;40:525–43.
[33] Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, et al. The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Res 2009;168:242–9.
[34] Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54:1063–70.
[35] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134:9–21.
[36] Belouchrani A, AbedMeraim K, Cardoso JF, Moulines E. A blind source separation technique using second-order statistics. IEEE Trans Signal Process 1997;45:434–44.
[37] Romero S, Mananas MA, Barbanoj MJ. A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: a simulation case. Comput Biol Med 2008;38:348–60.
[38] Ng SC, Raveendran P. Enhanced mu rhythm extraction using blind source separation and wavelet transform. IEEE Trans Biomed Eng 2009;56:2024–34.

[39] Shoker L, Sanei S, Chambers J. Artifact removal from electroencephalograms using a hybrid BSS–SVM algorithm. IEEE Signal Process Lett 2005;12:721–4.
[40] Gómez-Herrero G, De Clercq W, Anwar H, Kara O, Egiazarian K, Van Huffel S, et al. Automatic removal of ocular artifacts in the EEG without a reference EOG channel. In: Proc 7th Nordic Signal Process Symp (NORSIG). 2006. p. 130–3.
[41] Manly BFJ. Randomization, bootstrap, and Monte Carlo methods in biology. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC; 2007.
[42] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 1995:289–300.
[43] Lundqvist LO. Facial EMG reactions to facial expressions: a case of facial emotional contagion? Scand J Psychol 1995;36:130–41.
[44] Dimberg U, Thunberg M. Rapid facial reactions to emotional facial expressions. Scand J Psychol 1998;39:39–45.
[45] Sonnby-Borgstrom M, Jonsson P. Models-of-self and models-of-others as related to facial muscle reactions at different levels of cognitive control. Scand J Psychol 2003;44:141–51.
[46] Sonnby-Borgstrom M, Jonsson P, Svensson O. Gender differences in facial imitation and verbally reported emotional contagion from spontaneous to emotionally regulated processing levels. Scand J Psychol 2008;49: 111–22.
[47] Nishitani N, Hari R. Viewing lip forms: cortical dynamics. Neuron 2002;36:1211–20.
[48] Vaughan TM, Miner LA, McFarland DJ, Wolpaw JR. EEG-based communication: analysis of concurrent EMG activity. Electroencephalogr Clin Neurophysiol 1998;107:428–33.

Acknowledgements

Chapter 4, in full, is a reprint of the material as it appears in Behavioural Brain Research, 2012. Moore, Adrienne; Gorodnitsky, Irina; Pineda, Jaime. Adrienne Moore was the primary author and investigator of this paper.

CHAPTER 5

THE RECENT EVOLUTION OF MU SUPPRESSION RESEARCH

The mirroring properties of the mu rhythm of the EEG were published as early as 1954, when Bert and Gastaut observed changes in mu as subjects viewed a “moving picture” presentation of a boxing match (Gastaut & Bert, 1954). Rizzolatti et al., first reported mirror neurons in motor regions of macaque monkey cortex that fire both to performing and to observing goal directed hand movements in 1992 (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). However, it wasn’t until 2005 when a connection between these two domains was established and explored. EEG mu suppression was proposed and discussed by Pineda in 2005 as a means of measuring the putative human mirror neuron system (Pineda, 2005); that same year an experimental link between social deficits in ASD, the mirror neuron system, and mu suppression was published (Kilner & Lemon, 2013; Oberman et al., 2005), and subsequently interest in these topics took off.

The initial excitement regarding the promise of mu suppression research that followed early reports also inspired and was bolstered by widespread attention through the popular scientific press (Ramachandran & Oberman, 2006). Standing on the shoulders of these initial reports, a large proliferation of publications on these topics has emerged over the last ten years. The subset of literature focused on the infant mu rhythm alone is said to have more than doubled between 2011 and 2014 (Cuevas, Cannon, Yoo, & Fox, 2014). To illustrate, a PubMed search with the keywords “EEG mu suppression” yields 150 relevant search results, and 130 were published in 2009 or later (see Figure 1).

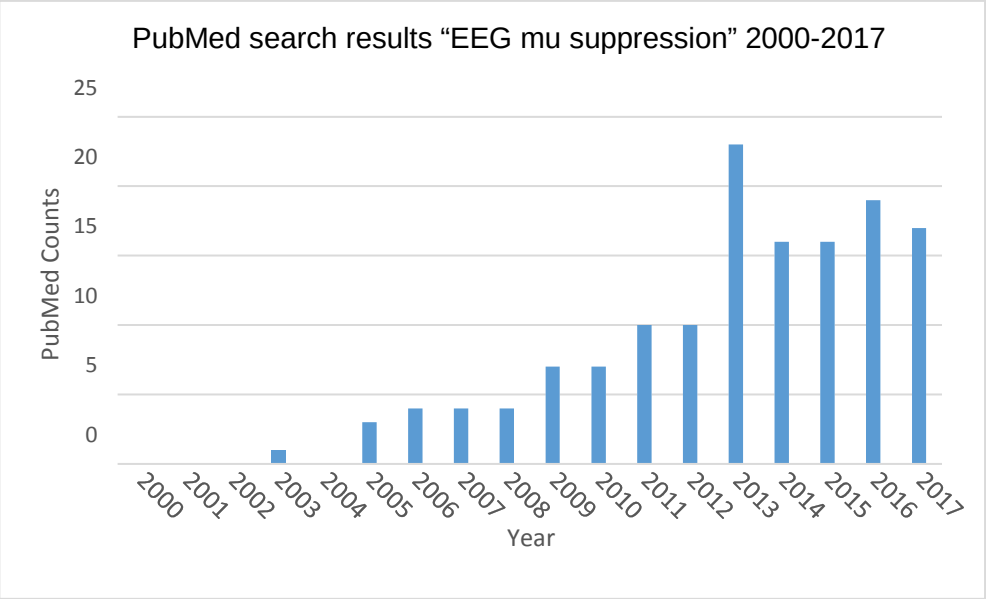

Figure 5.1: Increase in publications on EEG mu suppression

This early publication proliferation stage was naturally followed by a “pruning” phase, in order to guide the field of mu suppression research to a more mature stage. That is, the second wave of mu research reported some conflicting findings (Fan, Chen, Chen, Decety, & Cheng, 2014; Fan, Decety, Yang, Liu, & Cheng, 2010; Raymaekers, Wiersema, & Roeyers, 2009), and then an increased scrutiny of methodology began, in order to establish consistency within the rapidly growing mu suppression literature inspired by earlier work (Braadbaart, Williams, & Waiter, 2013; Dumas, Soussignan, Hugueville, Martinerie, & Nadel, 2014). Several critiques of methods were offered between 2013 and 2017, papers focused on establishing methodological standards for mu suppression research and identification of best practices to measuring mu suppression as an index of mirror neuron system activity (Tangwiriyasakul, Verhagen, van Putten, & Rutten, 2013), (Bowman et al., 2017; Cuevas et al., 2014), (H. M. Hobson & D. V. Bishop, 2017; Hobson & Bishop, 2016; H. M. Hobson & D. V. M. Bishop, 2017).

In light of these developments, this chapter will have two objectives: part one will describe the rationale for the methodological choices made for the 2012 paper reported in the previous chapter; and part two will summarize important publications with a methodological focus that were written after the 2012 publication of the previous chapter, regarding use of the mu rhythm to index the mirror neuron system.

Part 1: Methodological Considerations for Chapter 4

Blind Source Separation. ICA with source localization is used fairly often to study mu suppression in EEG data of late (Liao, Acar, Makeig, & Deak, 2015; Yin, Liu, & Ding, 2016), but was not the norm at the time Chapter 4 was published. Two considerations primarily motivated the application of a Blind Source Separation (BSS) technique to the EEG data presented in Chapter 4 in order to attempt to isolate the mu rhythm components, rather than using a more traditional analysis of electrode channel data. The first is given by Pfurtscheller and colleagues in a 1994 paper (Pfurtscheller & Neuper, 1994). Emphasizing the somatotopic organization of sensorimotor cortex, they refer to electrode sites C3 and C4 as positioned above and indexing a "hand area mu rhythm" specifically, and note that a face/tongue area mu rhythm exists as well. They also point out a "focal ERD/surround ERS" structure to the mu rhythm, and show that while a mu event-related desynchronization (ERD) indicating cortical activation is occurring at these hand electrode sites during a hand movement task, an event-related synchronization (ERS) can be recorded over neighboring cortical areas simultaneously and in the same frequency band. During a foot movement task, these neighboring electrodes reveal an ERD response, while the "hand electrodes" C3 and C4 reveal an ERS. Further, they show that the same focal/surround architecture applies to face compared to hand related activity, both movement and imagined movement. Mental imagery of orofacial movements (especially tongue movement and lip movement) produces an ERS response at electrodes C3

and C4, pronounced enough to be consistently identified in single trials (Pfurtscheller, Brunner, Schlogl, & Lopes da Silva, 2006; Spiegler, Graimann, & Pfurtscheller, 2004).

Consistent with this, in a preliminary channel analysis of this study's dataset indicated increased mu frequency power at electrodes C3/C4 found in face observation conditions relative to a control (Moore & Pineda 2007, conference poster). By this reasoning, the increase in power found at C3/C4 points to a possible decrease in power (or ERD) elsewhere, which is not directly indexed by any sensor electrodes in the sparse (13 electrode) array used for data collection in this study. Therefore a means of getting at the underlying neural generators for this face simulation cortical activation, such as BSS, was particularly desirable for this study's data analysis.

Preliminary application of the SOBI BSS algorithm for the purpose of routine artifact removal led to the second motivation for use of BSS on this dataset. The application of the BSS algorithm to this study's preprocessed data showed that after removal of electrooculographic (EOG) and electromyographic (EMG) components, the signals at both 10 Hz and 20 Hz at the central electrodes which are typically measured as the mu rhythm were the result of a mixture of various signal sources from several parts of the brain. Table 1 shows the degree of mixture, specifically the number of signal components contributing to the low and high mu frequency bands at central electrodes for each subject. The SOBI method estimates that there are a minimum of 2 generators in all cases, and sometimes more than 5 signal sources, contributing significantly (at least 10%) to the signal mixture commonly referred to as the mu rhythm. This evidence that 8-13 Hz power at C3 and C4 is not simply the mu rhythm but can be in fact a mixture of various brain sources is the second consideration pointing to the conclusion that use of a technique like BSS to un-mix these signals and isolate the mu rhythm would be beneficial.

Table 5.1: Number of SOBI components contributing at least 10% to mu EEG signals. The following chart shows for each subject the number of SOBI derived components that contributed at least 10% of the mu rhythm signal, that is, the EEG recorded from the lateral central electrodes (C3 and C4) within the low and high mu frequency bands at 10 and 20 Hz, after removal of visible artifacts. Data were analyzed using the EEGLAB 'component spectra and map' popspectopo function applied across the entire 3500 ms trial window and sampling 50% of the data per subject per condition. (Ss = subject number)

Ss:	C3 -10 Hz,	#:	C3 - 20 Hz,	#:	C4 - 10 Hz,:	#:	C4 - 20 Hz,	#:
1		2		4		4		5+
2		3		4		2		5+
3		2		4		3		3
4		4		4		2		3
5		3		4		3		4
6		3		4		4		4
7		4		3		2		2
8		2		5+		4		4
9		3		4		4		5+
10		5+		4		4		5+
11		4		5+		4		3
12		5+		3		2		4
13		2		3		3		3
14		2		5+		3		5+
15		2		3		4		3
16		3		4		3		5+
17		4		4		4		4
18		3		4		5+		4
19		3		4		3		5+
20		2		3		3		4
21		3		4		3		4
22		4		5+		4		3
23		4		3		4		3
24		3		4		2		4
25		2		3		3		2

Baseline selection. An additional methodological departure taken for our 2012 paper was with regard to choice of baseline. A number of early mu suppression studies included an extended baseline condition, but not a pre-stimulus baseline time period (Bernier, Dawson, Webb, & Murias, 2007; Oberman, Pineda, & Ramachandran, 2007; Oberman, Ramachandran, & Pineda, 2008). Typically mu power was averaged across each condition, then the

experimental and baseline conditions' power were put into a log-transformed mu suppression ratio, where 0 indicated the same amount of mu power in both conditions. Instead we used an event related spectral perturbations (ERD/ERS) design, with a within-trial pre-stimulus baseline. If the time course of the mu response involves both rising and falling power, averaging mu suppression across the conditions rather than using an event-related approach would not reveal this response, but ERSP could. Additionally, this approach has the advantage of testing not just whether power is greater or less in the experimental than in the control condition, but also whether or not stimulus related changes in spectral power (event related synchronizations or event related desynchronizations) occur in the control conditions, which was not usually examined.

Part 2: Toward Best Practices in EEG Mu Suppression Research

Intersubject variability and importance of baseline. Tangwiriyasakul et al. (2013) offer a systematic experimental investigation of what baseline condition is best for measuring relative event related mu suppression. The presented 5 sorts of baselines, compared to motor imagery: a gently moving flower, one bouncing ball, two moving balls, a static image of a hand, or a static black and white grid. They found that most but not all subjects had a discernable mu rhythm (11% were "mu absent"), and most but not all had mu suppression during mental imagery compared to at least one baseline (22% had no suppression in the mental imagery condition relative to any baseline). So 67% (12 subjects) confirmed mu is suppressed during mental imagery, at least relative to some baseline condition, when observing a hand opening and closing while instructed to imagine making the motion themselves. However, there was high inter-subject variability in which baseline was successful. Although the authors had tried a broad variety of baselines (both static and dynamic, complex and simple, biological and non-biological) they simply concluded that the baseline is important, and that in most subjects, mu

suppression during mental imagery will not be seen consistently across baselines, but only relative to a subset of plausible baselines. They were not able to conclude one of their tested baselines was better than another, or to recommend a particular type of baseline image, due to high intersubject variability in responses to baselines.

Event-related designs. Cuevas et al. (2014), in response to the steep increase in publication of infant EEG mu suppression studies, wrote a review article that both summarized the literature and proposed methodological guidelines to facilitate comparisons between studies of mu reactivity in typically developing infants. These authors also describe baseline considerations as their first and foremost concern. They recommend as a best practice for selection of baseline borrowing from an EEG study of infant macaque monkeys to facial gestures (Ferrari et al., 2012) which found suppressed 5-6 Hz EEG activity to producing and observing facial gestures, but not to observing non-biological stimuli, a study which they regard as particularly rigorous. The study's important attributes were including multiple baselines when the experimental condition is a dynamic, biological image, including one which is the static version of the experimental video stimulus, and one which preserves the movement but depicts a non-biological object in motion. This seems wise in principle, but they do not address inter-subject variability of responses to these sorts of baselines, which were reported (including static hand relative to moving hand, and moving, non-biological bouncing balls) by Tangwiriyasakul above.

Another baseline consideration, one not anticipated by Tangwiriyasakul et al., relates to timing of baselines. Cuevas et al., conclude the best method may be use of a "true event-related design", that is, one with discrete, short trials and in comparison a pre-stimulus baseline segment preceding each trial (as opposed to a lengthy baseline condition which is segmented off-line and averaged). This advice applies to both the baseline and the stimulus duration, that

is, Cuevas et al., also recommends using multiple, short (e.g. 1-3 seconds) stimulus presentations, in temporally close proximity to a particular action (e.g. a hand closing once), rather than a longer stimulus duration (e.g. tens of seconds covering various movements).

Beyond baselines. Cuevas et al., contribute several additional recommendations toward construction of standardized methods. They advise including both an "execute" (production of movement) condition and an "observe" condition which is perceptual and/or motor imagery and therefore relates to mirroring. This seems important particularly to developmental research using mu suppression, because the infant EEG is not measured from 8-13 Hz, it matures to that frequency band across development, and the execute condition allows the researcher to identify the appropriate frequency band for study at different points in development. Finally, Cuevas et al., devote substantial discussion to the recommendation of reporting EEG changes beyond the central sites. They state the occipital region may be the most important non-central site to include because the alpha rhythm is measured there. Alpha is found in the same 8-13 Hz frequency band as mu and is very sensitive to changes in attentional state, therefore it is a potential confound in mu suppression studies.

These are the primary recommendations made by Cuevas et al., while their additional recommendations are about more general research practices (e.g. using live observation trials if possible for maximum ecological validity, reporting outliers found and minimum amounts of data included). Although focused on infant research, with the possible exception of the necessity of including an action execution as well as an action observation/imagery condition which facilitates identifying the developmentally appropriate frequency to measure, these recommendations apply well to adult research also.

Surprising meta-analysis results. Fox et al. (2016) present a meta-analysis of 85 mu EEG studies encompassing 1707 participants that all infer human mirroring system activity,

and additionally make suggestions for improving methods used in this sort of research. Overall, they confirm significant effect sizes for mu suppression during both action execution (Cohen's d=.46) and action observation (Cohen's d=.31) across the 85 studies. The latter finding should confirm that the mu rhythm does in fact reflect recruitment of sensorimotor cortex during perceptual processing of actions, or mirroring. They also include a careful analysis of whether publication bias and non-publication of insignificant results are impacting their result and conclude that despite the likelihood of some influence of publication bias, their effect remains significant after accounting for this. However, they also report somewhat surprising results related to the other considerations they address, namely, theoretically relevant moderators, methodological moderators, and topographic specificity of effects.

Theoretical moderators not as predicted. The theoretical moderators considered were whether or not the action observed was biological motion, and whether or not the action (executed or observed) was object oriented. The former is predicted to be a significant moderator based on the theory that mu suppression reflects a simulation of another biological agent's body movement; the latter is predicted to be a significant moderator because mirror neurons in macaques fire to object directed but not non-object directed actions. Interestingly, neither of these theoretically relevant moderators was significant (in fact the direction of results associated biological stimuli with smaller effects (Cohen's d=.30) than non-biological stimuli (Cohen's d=.51)).

Methodological moderators not as predicted. The methodological factors considered were two aspects of sample composition (age (0-4, 5-18, >18 years) and gender) plus type of baseline used. Types of baseline were static or dynamic, biological or non-biological, live or video. Surprisingly, a significant effect of gender was found, with studies of males reporting significantly larger mu suppression effects (Cohen's d=.38) than studies of females (Cohen's

d=.27). This was not expected given that multiple studies that set out to examine gender differences in mu suppression have concluded females exhibit stronger mu suppression in mirroring conditions (Cheng et al., 2008; Yang, Decety, Lee, Chen, & Cheng, 2009). Neither age nor type of baseline was a significant moderator of effects. This null finding with regard to baseline doesn't contradict Tangwiriyasakul's prior assertion that choice of baseline is important because many subjects will appear mu suppressed compared to one baseline and not to another. Rather, it is similar to their assertion that no particular type of baseline was found consistently effective, because there was high inter-subject variability in baseline responses.

Confounding alpha activity. The findings of Fox et al., regarding their third consideration, topographic specificity, were particularly striking. The authors did confirm topographic specificity for action execution, showing that frontal and occipital scalp locations had significantly smaller 8-13 Hz suppression effects than central locations. However, the authors were not able to confirm topographic specificity of suppression for action observation. In the subset of studies that reported data from non-central electrodes, mu suppression was no greater than suppression in the same frequency band at occipital and other scalp locations.

This lack of specificity to central electrodes casts some doubt on claims that sensorimotor mirroring areas, including those homologous to the macaque mirror neuron system, are necessarily the source of the significant suppression effects reported at central electrodes during many EEG studies of action observation. This also might explain the null findings of the meta-analysis with regard to theoretically relevant moderators, i.e. whether or not the action was object directed or performed by a biological agent. Both these moderators would only be expected to impact results if 8-13 Hz EEG suppression at electrodes C3 and C4 specifically reflects sensorimotor cortex activation, the source of mirroring or action simulation.

As a suggestion for future research, the authors state that attentional confounds, probably stemming from alpha rhythm contamination of the mu signal, must be better controlled, or alternately source localization could be used to confirm the signal from central electrodes measured as the mu rhythm really does reflect a sensorimotor cortex response.

Attentional confounds are mentioned in the reviewed papers, that is, some attempts were being made to control for attention, sometimes by discarding portions of data most likely to be impacted by attention, or by engaging the subjects in a task that forces them to remain attentive such as counting (Oberman et al., 2008), or for infants choosing stimuli they think the subjects find interesting (van Elk, van Schie, Hunnius, Vesper, & Bekkering, 2008). The problem is that there's often no way to confirm that these attempts were successful at equalizing attentional salience across baseline and experimental conditions for each or most subjects, or in some cases it is evident that they were not successful because posterior alpha is desynchronized along with mu (Warreyn et al., 2013) (Paulus, Hunnius, & Bekkering, 2013) (Perry, Troje, & Bentin, 2010).

Hobson and Bishop (2016) responded to the previous papers by conducting a study on 61 neurotypical adults (the largest mu suppression study to date) specifically with the aim of investigating the validity of mu suppression as a measure of the human mirror neuron system. They focused on the impact of different baseline measures, and comparison of activity at both central and occipital electrodes which could reveal an attentional confound, as well as mirroring in the beta (13-35 Hz) frequency band in addition to the 8-13 Hz band, as this frequency band has been reported to reflect mirroring in some experiments as well. Stimuli presented were two biological motion observation conditions, presenting a hand action both with and without an object (a pencil), and a baseline condition presenting a video of a dynamic, geometric "kaleidoscope" design (very similar to the geometric stimulus used in Chapters 2 and 3 of this

dissertation). A short (8 second) and a long (80 second) baseline resting condition were also included, wherein subjects looked at a blank screen without moving. A third baseline option was created within each trial by presenting the videos as static images for several seconds before they became dynamic, and comparing the mu power pre- and post- video onset. An "execute" condition was also included, which instructed subjects to tap index finger and thumb together for 40 seconds. EMG was simultaneously recorded from the arm muscle that performs finger extension, to see whether automatic imitation occurred in participants, and to remove individual trials contaminated with movement artifacts.

Attempt to clarify with a large study. Hobson and Bishop (2016) calculated mu suppression for the execute condition and each of the three observe conditions (hand + object, hand no object, kaleidoscope) relative to each of the three baselines (long, short, and pre-stimulus/within trial) in a sample of 61 adults. They identified the result that would successfully confirm mu suppression as a valid indicator of mirroring: an interaction between electrode site and condition such that the difference between kaleidoscope and biological motion observation was greater at central mu than occipital alpha sites. This would both show that suppression was specific to central, sensorimotor sites, and clarify whether the result is sensitive to baseline selection, in a large, well powered study.

Their results for the beta frequency band did not match this pattern or offer validating evidence of mirroring at central electrodes and higher frequencies. Their results for the traditional 8-13 Hz mu showed that only one of the three baseline methods used revealed a main effect of electrode site and an interaction between electrode and observed stimulus. Mu suppression calculated relative to a pre-stimulus/within trial baseline was significantly greater when viewing hand movement than when viewing kaleidoscope movement, at central but not

occipital electrodes (while kaleidoscope induced greater 8-13 Hz alpha suppression at occipital electrodes).

This is a reassuring finding indicating that with the correct selection of baseline and rigorous attention to experimental design, a mirroring response that is not confounded with occipital alpha generated attentional responding can be measured. Nevertheless, the authors express skepticism regarding the utility of mu suppression as an index of the mirroring system. They point out that only in the hand interacting with an object and not the hand movement alone condition were results significant, plus at the individual subject level 16-21% of their neurotypical participants did not show mu suppression during hand + object observation, and they refer to mu suppression mirroring effects as "weak and unreliable and easily confounded with alpha suppression." Fox's group rapidly responded in a 2017 article with a more optimistic interpretation of the utility of measuring mu suppression to index mirroring titled, "The mu-rhythm can mirror: Insights from experimental design, and looking past the controversy". A few months later, Hobson and Bishop responded with a targeted brief report, and with a lengthy counter argument as a second article, "The interpretation of mu suppression as mirror neuron activity: past, present and future".

Key points of agreement. It would be redundant to individually detail these most recent papers as they cover a lot of the same ground regarding different schools of thought on mu suppression and mirroring. But while Fox et al., and Hobson and Bishop remain in opposing roles as the self-described optimists and skeptics respectively during this dialogue, they agree on key points. For instance, they agree 1) that mu suppression reflects mirroring sometimes, 2) that the key to indexing mirroring with mu suppression is distinguishing sensorimotor mu from occipital alpha, and mirroring effects from visual attention effects, 3) that an event-related pre-stimulus baseline within each trial is most effective, and 4) that the most

rigorous experimental designs for mu suppression research include both mirroring condition/s and an “execute” condition as a positive control.

Independent Component Analysis. An additional new argument from the Fox group in their latest publication points out that requiring that occipital electrode 8-13 Hz power not be suppressed, or be significantly less suppressed, during action observation compared to central electrode 8-13 Hz power in order to conclude mirroring occurred isn’t necessarily the best approach either. Alpha suppression reflecting visual attentional engagement at occipital electrodes as well as mu suppression reflecting mirroring at central electrodes in response to the same visual stimulus is entirely plausible, they aren’t mutually exclusive. This leads Hobson and Bishop in their most recent paper to add a discussion of the use of ICA (or another blind source separation tool) to dissociate and isolate mu and alpha components, and presenting results in component space rather than channel space. This should obviate the well-known low spatial resolution issues of EEG, which are due to volume conduction that creates the problem of the alpha rhythm appearing at central electrodes in the first place. It is noted that this approach is relatively unexplored and may prove to be key to future mu suppression research, if it can rule out confounds that have complicated the field in the past.

Pre-registration. Finally, Hobson and Bishop (2016) argue convincingly for the practice of pre-registration of research plans, so that hypotheses, analyses and what will be reported are determined before data collection, to avoid bias toward reporting positive findings (H. M. Hobson & D. V. Bishop, 2017). In that spirit, I will mention some null results that were found in the study presented in Chapter 4. Three measurements were taken that could be used to index individual differences in subjects’ empathy, which could then be tested for correlation with amount of mu suppression. Subjects rated on a 1-5 scale how well they believe they empathized with each face after each trial; they reported their mood using the Positive

and Negative Affect Scale questionnaire twice, after the negative and positive empathy conditions, and a difference score was calculated to see whether there mood changed in the direction empathy would predict; and they completed the Balanced Emotional Empathy Scale, a trait measurement of differences in responsiveness to the emotional experiences of others.

Chapter 4 null findings. None of these empathy measures were significantly correlated with amount of mu suppression in response to emotional faces (a fact I considered uninteresting at the time and failed to discuss in the publication). Overall, almost every area of literature reviewed for the introduction to this dissertation (including facial mimicry, emotional contagion, mirroring revealed by fMRI, mirroring revealed by EEG, mirroring in ASD, eye vs mouth looking in ASD) involved a large number of studies that were later contradicted and which failed to replicate, a problem to which I don't wish to contribute. As others have noted, reporting null results and shifting the scientific culture toward 'you can publish if the study design is rigorous' (determined prior to examination of any data) rather than 'you can publish if results are significant' (which motivates finding your way to a statistically significant result with every dataset) might be a great means of lessening false positives that later fail to replicate and improving social neuroscience (Forstmeier, Wagenmakers, & Parker, 2017) .

Comments on Chapter 4 results. Although the studies summarized here were all published later than "EEG mu component responses to viewing emotional faces", we did anticipate and attempt to address the main problems raised in the recent critiques of mu suppression as an index of mirror system activity: how to isolate a mu (not alpha) signal from sensorimotor cortex, and what baseline to select relative to which to define mu suppression. However, our "control" condition of viewing buildings did show a mu suppression pattern of results, though not as robust as that of face conditions, even with carefully selected mu components based on blind source separation and a brief baseline calculated within the same

trial as the mu suppression. Specifically, 50% of left hemisphere mu components and 47% of right hemisphere mu components were significantly desynchronized relative to baseline when viewing buildings; viewing faces elicited a significant desynchronization in 81% and 77% of left and right components.

There are at least two possible explanations for this finding. A task was given to subjects to encourage them to remain engaged (button press rating of attractiveness of faces and buildings in the two non-empathy conditions; trying to empathize and then rating how well they empathized in the empathy to faces condition). Anticipation of the button press, essentially mentally rehearsing it, causing mu suppression due to mental simulation of a hand action can't be ruled out as a possible cause of mu suppression to buildings. Measuring EMG from hand electrodes during the experiment would address this limitation. Alternately attentional engagement may still be influencing mu component responses across conditions, even with the use of blind source separation. That is, viewing visual noise may be no more engaging than the pre-stimulus fixation cross baseline, while viewing and rating buildings may be slightly more engaging, and viewing and rating faces consistently more engaging. Plus, either SOBI blind source separation may not have adequate separated mu rhythm from alpha activity, or the mu rhythm even when perfectly separated from the alpha rhythm may be sensitive to attentional salience.

Comments on Chapter 4 methods. To address this, alpha components could be identified using SOBI blind source separation, and either removed from data as a preprocessing step, or compared across conditions as a control if a different pattern of results was consistently found from alpha components in response to stimuli. However, the source localization of mu components utilizes the fortuitous location of mu source generators, which lie fairly superficially in the cortex such that they tend to be well resolved by source localization

algorithms, something that may not be true of alpha source generators. Therefore, it remains to be seen whether the method used here (with the criteria for identification of 8-13Hz frequency band of peak power, confirmed by scalp map projections and dipole localization) would actually be as effective at identifying alpha as mu components.

Conclusion

If I were doing this experiment today I would do many things the same way, however I would 1) add an “execute” condition (instructing subjects to produce the facial expression presented), which might make the separation of the mu component more robust; 2) try to isolate and identify the alpha components as well as the mu components to prove they are distinct in function, which might entail using a greater number of electrodes to better isolate alpha components; 3) consider ICA instead of SOBI for blind source separation as its usage seems more standard and prevalent; and 4) prior to considering faces, first try to replicate Hobson and Bishop 2016’s findings (of mu suppression when observing hand actions with a pencil) using blind source separation. When both the method’s efficacy (e.g. pros and cons of using either ICA or SOBI or traditional channel space to measure mu suppression) and the hypothesis (whether mu suppression occurs when observing emotional faces and is related to empathy) are both unknown and being tested simultaneously, as was the case in “EEG mu component responses to viewing emotional faces,” an experiment is intrinsically exploratory. The past few years have been an exciting time for mu suppression research. Despite an upsurge in number of publications, many questions, including methodological questions, remain unanswered or only partially answered, even with regard to neurotypical adult subjects, arguably the most basic starting point.

Proceeding cautiously through the mu suppression literature (i.e. considering studies using within trial baselines, both execute and observe conditions, either reporting occipital as

well as central electrode results or using ICA and source localization to separate mu and alpha, and reporting some control condition that does not suppress mu), there are a number of interesting results from studies of mu suppression in infants. During observation of reaching and grasping at 12 months, an ERD response can be observed in the same frequency band as for executing the action, around 7 Hz (Thorpe, Cannon, & Fox, 2016). While the topography of the response does not include occipital electrodes, it does occur in a diffuse region across frontal central and parietal electrodes at 12 months, becoming more focal and adult-like in 4 year olds. In additional studies, the development of this infant mu suppression during hand action observation has been associated with increases in the infant's own grasping skills (Cannon et al., 2016; Yoo, Cannon, Thorpe, & Fox, 2016). Regarding perception of dynamic positive and negative facial expressions and emotionally neutral mouth opening faces, in 30 month old infants, mu ERD was shown during face observation, right lateralized for emotional faces and bilaterally for the mouth opening face (Rayson, Bonaiuto, Ferrari, & Murray, 2016). This finding reaffirms the fundamental claim of Chapter 4, that mirroring of emotional facial expressions is reflected in mu suppression, measured as event related desynchronization that is focused above sensorimotor cortex.

In conclusion, it appears from Chapters 1 and 4 that there is ample evidence that there is a human mirroring system, and from Chapter 5 that mu suppression can reflect it given rigorous methodological practices. What remains unclear is whether mu suppression is reliable enough at the individual subject level to serve as a biomarker, an objectively measurable characteristic that indicates presence or absence of a pathogenic condition (Strimbu & Tavel, 2010). Because the questions that might be addressed by mu suppression research related to knowledge of other minds, empathy, the organization of social cognition in the brain, and the origin of the social impairments in ASD are so fundamental and intriguing, no doubt the upcoming years will yield exciting new research exploring these topics

BIBLIOGRAPHY

Bernier, R., Dawson, G., Webb, S., & Murias, M. (2007). EEG mu rhythm and imitation impairments in individuals with autism spectrum disorder. *Brain Cogn, 64*(3), 228-237. doi:10.1016/j.bandc.2007.03.004

Bowman, L. C., Bakermans-Kranenburg, M. J., Yoo, K. H., Cannon, E. N., Vanderwert, R. E., Ferrari, P. F., . . . Fox, N. A. (2017). The mu-rhythm can mirror: Insights from experimental design, and looking past the controversy. *Cortex, 96*, 121-125. doi:10.1016/j.cortex.2017.03.025

Braadbaart, L., Williams, J. H., & Waiter, G. D. (2013). Do mirror neuron areas mediate mu rhythm suppression during imitation and action observation? *Int J Psychophysiol, 89*(1), 99-105. doi:10.1016/j.ijpsycho.2013.05.019

Cannon, E. N., Simpson, E. A., Fox, N. A., Vanderwert, R. E., Woodward, A. L., & Ferrari, P. F. (2016).

Relations between infants' emerging reach-grasp competence and event-related desynchronization in EEG. *Dev Sci, 19*(1), 50-62. doi:10.1111/desc.12295

Cheng, Y., Lee, P. L., Yang, C. Y., Lin, C. P., Hung, D., & Decety, J. (2008). Gender differences in the mu rhythm of the human mirror-neuron system. *PLoS One, 3*(5), e2113. doi:10.1371/journal.pone.0002113

Cuevas, K., Cannon, E. N., Yoo, K., & Fox, N. A. (2014). The Infant EEG Mu Rhythm: Methodological Considerations and Best Practices. *Dev Rev, 34*(1), 26-43. doi:10.1016/j.dr.2013.12.001

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp Brain Res, 91*(1), 176-180.

Dumas, G., Soussignan, R., Hugueville, L., Martinerie, J., & Nadel, J. (2014). Revisiting mu suppression in autism spectrum disorder. *Brain Res, 1585*, 108-119. doi:10.1016/j.brainres.2014.08.035

Fan, Y. T., Chen, C., Chen, S. C., Decety, J., & Cheng, Y. (2014). Empathic arousal and social understanding in individuals with autism: evidence from fMRI and ERP measurements. *Soc Cogn Affect Neurosci, 9*(8), 1203-1213. doi:10.1093/scan/nst101

Fan, Y. T., Decety, J., Yang, C. Y., Liu, J. L., & Cheng, Y. (2010). Unbroken mirror neurons in autism spectrum disorders. *J Child Psychol Psychiatry, 51*(9), 981-988. doi:10.1111/j.1469-7610.2010.02269.x

Ferrari, P. F., Vanderwert, R. E., Paukner, A., Bower, S., Suomi, S. J., & Fox, N. A. (2012). Distinct EEG amplitude suppression to facial gestures as evidence for a mirror mechanism in newborn monkeys. *J Cogn Neurosci, 24*(5), 1165-1172. doi:10.1162/jocn_a_00198

Forstmeier, W., Wagenmakers, E. J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings - a practical guide. *Biol Rev Camb Philos Soc, 92*(4), 1941-1968. doi:10.1111/brv.12315

Gastaut, H. J., & Bert, J. (1954). EEG changes during cinematographic presentation; moving picture activation of the EEG. *Electroencephalogr Clin Neurophysiol, 6*(3), 433-444.

Hobson, H. M., & Bishop, D. V. (2017). The interpretation of mu suppression as an index of mirror neuron activity: past, present and future. *R Soc Open Sci, 4*(3), 160662. doi:10.1098/rsos.160662

Hobson, H. M., & Bishop, D. V. M. (2016). Mu suppression - A good measure of the human mirror neuron system? *Cortex, 82*, 290-310. doi:10.1016/j.cortex.2016.03.019

Hobson, H. M., & Bishop, D. V. M. (2017). Reply to Bowman et al.: Building the foundations for moving mu suppression research forward. *Cortex, 96*, 126-128. doi:10.1016/j.cortex.2017.06.012

Kilner, J. M., & Lemon, R. N. (2013). What we know currently about mirror neurons. *Curr Biol, 23*(23), R1057-1062. doi:10.1016/j.cub.2013.10.051

Liao, Y., Acar, Z. A., Makeig, S., & Deak, G. (2015). EEG imaging of toddlers during dyadic turn-taking: Mu-rhythm modulation while producing or observing social actions. *Neuroimage, 112*, 52-60. doi:10.1016/j.neuroimage.2015.02.055

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Brain Res Cogn Brain Res, 24*(2), 190-198. doi:10.1016/j.cogbrainres.2005.01.014

Oberman, L. M., Pineda, J. A., & Ramachandran, V. S. (2007). The human mirror neuron system: a link between action observation and social skills. *Soc Cogn Affect Neurosci, 2*(1), 62-66. doi:10.1093/scan/nsl022

Oberman, L. M., Ramachandran, V. S., & Pineda, J. A. (2008). Modulation of mu suppression in children with autism spectrum disorders in response to familiar or unfamiliar stimuli: the mirror neuron hypothesis. *Neuropsychologia, 46*(5), 1558-1565. doi:10.1016/j.neuropsychologia.2008.01.010

Paulus, M., Hunnius, S., & Bekkering, H. (2013). Neurocognitive mechanisms underlying social learning in infancy: infants' neural processing of the effects of others' actions. *Soc Cogn Affect Neurosci, 8*(7), 774-779. doi:10.1093/scan/nss065

Perry, A., Troje, N. F., & Bentin, S. (2010). Exploring motor system contributions to the perception of social information: Evidence from EEG activity in the mu/alpha frequency range. *Soc Neurosci, 5*(3), 272-284. doi:10.1080/17470910903395767

Pfurtscheller, G., Brunner, C., Schlogl, A., & Lopes da Silva, F. H. (2006). Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *Neuroimage, 31*(1), 153-159. doi:S1053-8119(05)02514-0 [pii]10.1016/j.neuroimage .2005.12.003

Pfurtscheller, G., & Neuper, C. (1994). Event-related synchronization of mu rhythm in the EEG over the cortical hand area in man. *Neurosci Lett, 174*(1), 93-96.

Pineda, J. A. (2005). The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing". *Brain Res Brain Res Rev, 50*(1), 57-68. doi:10.1016/j.brainresrev.2005.04.005

Ramachandran, V. S., & Oberman, L. M. (2006). Broken mirrors: a theory of autism. *Sci Am, 295*(5), 62-69.

Raymaekers, R., Wiersema, J. R., & Roeyers, H. (2009). EEG study of the mirror neuron system in children with high functioning autism. *Brain Res, 1304*, 113-121. doi:10.1016/j.brainres.2009.09.068

Rayson, H., Bonaiuto, J. J., Ferrari, P. F., & Murray, L. (2016). Mu desynchronization during observation and execution of facial expressions in 30-month-old children. *Dev Cogn Neurosci, 19*, 279-287. doi:10.1016/j.dcn.2016.05.003

Spiegler, A., Graimann, B., & Pfurtscheller, G. (2004). Phase coupling between different motor areas during tongue-movement imagery. *Neurosci Lett, 369*(1), 50-54. doi:10.1016/j.neulet.2004.07.054

Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Curr Opin HIV AIDS, 5*(6), 463-466. doi:10.1097/COH.0b013e32833ed177

Tangwiriyasakul, C., Verhagen, R., van Putten, M. J., & Rutten, W. L. (2013). Importance of baseline in event-related desynchronization during a combination task of motor imagery and motor observation. *J Neural Eng, 10*(2), 026009. doi:10.1088/1741-2560/10/2/026009

Thorpe, S. G., Cannon, E. N., & Fox, N. A. (2016). Spectral and source structural development of mu and alpha rhythms from infancy through adulthood. *Clin Neurophysiol, 127*(1), 254-269. doi:10.1016/j.clinph.2015.03.004

van Elk, M., van Schie, H. T., Hunnius, S., Vesper, C., & Bekkering, H. (2008). You'll never crawl alone: neurophysiological evidence for experience-dependent motor resonance in infancy. *Neuroimage, 43*(4), 808-814. doi:10.1016/j.neuroimage.2008.07.057

Warreyn, P., Ruysschaert, L., Wiersema, J. R., Handl, A., Pattyn, G., & Roeyers, H. (2013). Infants' mu suppression during the observation of real and mimicked goal-directed actions. *Dev Sci, 16*(2), 173-185. doi:10.1111/desc.12014

Yang, C. Y., Decety, J., Lee, S., Chen, C., & Cheng, Y. (2009). Gender differences in the mu rhythm during empathy for pain: an electroencephalographic study. *Brain Res, 1251*, 176-184. doi:10.1016/j.brainres.2008.11.062

Yin, S., Liu, Y., & Ding, M. (2016). Amplitude of Sensorimotor Mu Rhythm Is Correlated with BOLD from Multiple Brain Regions: A Simultaneous EEG-fMRI Study. *Front Hum Neurosci, 10*, 364. doi:10.3389/fnhum.2016.00364

Yoo, K. H., Cannon, E. N., Thorpe, S. G., & Fox, N. A. (2016). Desynchronization in EEG during perception of means-end actions and relations with infants' grasping skill. *Br J Dev Psychol, 34*(1), 24-37. doi:10.1111/bjdp.12115

CHAPTER SIX

SUMMARY AND CONCLUSION

Autism Spectrum Disorder is a major public health concern, with an estimated cost burden of $17,000 per child per year (Lavelle et al., 2014). Current practices for identifying children with ASD rely on fallible self-report questionnaires for screening, and expert clinician judgment, which is unavailable to a large number of people, for diagnosis (Filipek et al., 2000; Hayes, Ford, Rafeeque, & Russell, 2018; Towle & Patrick, 2016). Objective, standardized behavioral measures that are derived from and informed by neuroscience research, as well as direct neural measures, could provide a powerful complement to our existing suite of tools for meeting the public health challenge of ASD.

Chapter 2 has shown that the GeoPref eye tracking test for ASD identification is comparably effective for early detection of both males and females with ASD (with ROC AUC's of .72 and .75 respectively). Also, those with eye tracking scores at the extremes (which we refer to as the GeoPref and SocPref subtypes) do have similar relationships to clinical phenotype measures in females as in males. That is, females with increased social preference according to eye tracking have a strong advantage over the GeoPref subgroup in verbal developmental abilities. Further, the SocPref subtype is significantly more prevalent among non-ASD females than among non-ASD males, supporting subtle sex differences in neurotypical social development. Finally, considering a set of unaffected younger siblings to an older sibling with ASD, these high risk females do exhibit enhanced social attention on the GeoPref test, while high risk males do not. This social attention difference in those at risk for autism could point toward the underlying mechanisms that modulate the strikingly different vulnerability of males and females to ASD.

Chapter 3 has shown that a new Complex Social stimulus video, while increasing the frequency and duration of dyadic interactions and emotional faces depicted during eye

tracking, maintains the property of detecting individual ASD toddlers based on elevated time spent viewing geometric images (with slightly improved ROC AUC of .75 for the Complex Social test, compared to .73 for the original GeoPref test, averaged across sexes). Furthermore, the Complex Social test had the additional property of ruling out ASD in a subset of toddlers who spend the least time viewing geometric images (none of whom were diagnosed ASD). Moreover, combined use of the original and Complex Social GeoPref tests does improve sensitivity of these tools (from 23% to 35%) for detection of toddlers with ASD. Increasing the sensitivity while maintaining the high specificity of eye tracking tests to detect ASD by combining several into a test battery could potentially strongly impact the usability of these tools.

Chapter 4 has shown that mirroring of emotional facial expressions is measureable with EEG mu component suppression, using components extracted by blind source separation techniques Further, perception of positive (smiling) and negative (disgusted) facial expressions impact mu component suppression differently, particularly in the right hemisphere. However, individual differences in mu suppression when viewing emotional faces were not correlated with individual differences in trait empathy as predicted. Methodological best practices have been the focus of several important recent mu suppression publications (discussed in Chapter 5). These highlight the importance of selecting an appropriate baseline to measure mu suppression against, and eliminating attentional confounds due to the mixing of occipital alpha signals with the sensorimotor mu signal. Mirror neurons are now listed among the NIMH research domain criteria (RDoC) initiative constructs at the cellular unit of analysis for Reception of Social Communcation, and mu suppression is included in NIMH RDoC as well, at the circuit level for studying Perception and Understanding of Others (www.nimh.nih.gov/research-priorities/rdoc), indicating their potential to advance knowledge of the brain circuitry underlying neurotypical and ASD behavior (Insel, 2014).

Current “gold standard” tools for identifying toddlers with ASD involve a highly trained clinician observing complex behaviors (e.g. the ADOS presses a child to act out a doll’s bath time routine). These behaviors are fairly distant from and difficult to interpret in terms of current understanding of the organization of cognitive and neural systems. Eye tracking may be a particularly useful behavioral measure for autism research, because it lies at an intermediate level between the everyday activities where autism can reveal itself to clinicians, and underlying neurocognitive networks for perception, attention and social cognition believed to function atypically in ASD (Falck-Ytter, Bolte, & Gredeback, 2013). A neural measure, particularly if it were affordable and noninvasive like EEG mu suppression, would be an invaluable tool if it accurately identified individual infants and young toddlers with ASD or had other clinical applications, such as prognosis/stratification, or measurement of response to treatment.

However, there is a deeper problem, related to our frame of reference for determining what accurate identification of ASD means (Walter, 2013). Tom Insel (head of the NIMH from 2002 to 2015) has frequently asserted that behaviors in mental disorders including ASD are just symptoms, while their causes are the more important question, and their causes lie in neural circuits and biology (Insel, 2014; Insel & Cuthbert, 2015). Is the current state of ASD diagnosis moving close to a biological approach, something like that expressed by Insel? ASD diagnosis through the mainstream healthcare system in the U.S. relies on the Diagnostic and Statistical Manual of mental disorders. DSM definitions of ASD differ considerable from DSM-III to DSM-IV to DSM-5, for example the category PDD-NOS (persistent developmental delay – not otherwise specified) was eliminated, and the category ASD encompasses the remaining persistent developmental disorders from DSM-IV (Tanguay, 2011). However from a perspective like that of Insel, it is difficult to say that one definition is more correct than another, because none are based on identifying any biological abnormality (Insel & Cuthbert, 2015).

There certainly has been improvement in ASD diagnosis in recent decades, for example Cathy Lord's creation and development of the ADOS, which has been accepted as a gold standard diagnostic tool and is now widely used (Falkmer, Anderson, Falkmer, & Horlin, 2013; Lord, 1991). As a result, if you have access to the gold standard diagnostic practices, it is probably considerably more likely than it would otherwise be that if you go to two clinicians they will give you the same answer as to whether or not your young child has ASD. Additionally, the age at which a stable diagnosis of ASD can be given has decreased (Guthrie, Swineford, Nottke, & Wetherby, 2013). But these sorts of developments have more to do with greater internal consistency of clinician practices than correspondence between an ASD diagnosis and any particular etiology or biological difference.

To illustrate, Sally Ozonoff published an editorial perspective commenting on factors that lead to removal of PDD-NOS as a DSM-V category (Ozonoff, 2012). A large, 12 site study in 2011 showed that there was a gap between scores on the ADOS assessment and clinician judgments regarding who actually qualified for a diagnosis of PDD-NOS (clinicians interpreted and weighted the same assessment scores differently) (Lord et al., 2012). "The strongest predictor of diagnosis was what site made it, rather than any characteristic of the child. This is a clear sign that the PDD subtypes were just not working" (Ozonoff, 2012).

I thought this was fascinating because Ozonoff is arguing that changing the definition of the disorder is a solution to difficulties in identification of those with a behavioral disorder, which is not how we think about disorders that correspond to known biology. For example, if pediatricians can't look into an ear and say whether an ear infection is bacterial and should be prescribed an antibiotic, or is viral and should just be treated with pain management, no one thinks we should do away with the ear infection subtypes 'bacterial' and 'viral' so that all clinicians will agree. Yet this position is understandable and pragmatic. To quote Ozonoff (2012) again, "From my perspective as a practicing clinician, if these conditions cannot be

validly distinguished empirically and the labels are used inconsistently, then it is logically questionable (as well as patently unfair) to deny services to some while providing a full range of interventions to others."

The goals of clinicians have much to do with helping connect challenged families with resources, specifically with the most appropriate types of resources, and ideally in a manner that is equitable and fair (Lord & Jones, 2012). Perhaps validity in diagnosis of mental illness with those goals in mind is primarily about reliability, consistency or creating a system of practice that is effective (Mahjouri & Lord, 2012). This just seems quite distant from the perspective of, for example, Tom Insel who seems to believe correspondence between behavior and a biological substrate is the type of validity we should be seeking, where behaviors are just a proxy we are using until the underlying neurobiology is understood (Cuthbert & Insel, 2013).

There are also a number of concerns with access to ADOS trained clinicians, and reported long wait times between first signs of concern and actually receiving a diagnosis and then entering treatment (Vohra, Madhavan, Sambamoorthi, & St Peter, 2014). For this, eye tracking could be helpful, if it basically comes to the same conclusion as a clinician would but does so more quickly and less expensively. The eye tracking and EEG methods explored in this dissertation are also language-free, which could help remove barriers to accessing services (Zuckerman et al., 2013). But if we are serious about moving past behavioral definitions, we can't define success in terms of being as sensitive and as specific as the ADOS is in classifying to DSM categories. There's little reason to expect a strong biomarker to exist for a category that is not actually biological, rather we should expect biomarkers to "cut across current heterogeneous categories of mental disorders" (Casey, Oliveri, & Insel, 2014). We would therefore have to be willing to reorient research away from DSM categories, and what that would probably look like is poor sensitivity or poor specificity or both according to current

definitions (Kapur, Phillips, & Insel, 2012). But to abandon DSM would be to abandon mainstream healthcare systems, including insurance reimbursement, and determination of disability and service eligibility, at least in the short-term (Casey et al., 2013). These seem to me a very difficult set of problems to solve, but eminently worthwhile as well, as a truly new biomarker approach to ASD identification could lead to far better patient experiences. Perhaps equally compelling, these inquiries into brain and behavior position a few additional pieces of an old philosophical puzzle, that of how humans in their daily lives solve the problem of other minds.

BIBLIOGRAPHY

Casey, B. J., Craddock, N., Cuthbert, B. N., Hyman, S. E., Lee, F. S., & Ressler, K. J. (2013). DSM-5 and RDoC: progress in psychiatry research? *Nat Rev Neurosci, 14*(11), 810-814. doi:10.1038/nrn3621

Casey, B. J., Oliveri, M. E., & Insel, T. (2014). A neurodevelopmental perspective on the research domain criteria (RDoC) framework. *Biol Psychiatry, 76*(5), 350-353. doi:10.1016/j.biopsych.2014.01.006

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med, 11*, 126. doi:10.1186/1741-7015-11-126

Falck-Ytter, T., Bolte, S., & Gredeback, G. (2013). Eye tracking in early autism research. *J Neurodev Disord, 5*(1), 28. doi:10.1186/1866-1955-5-28

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: a systematic literature review. *Eur Child Adolesc Psychiatry, 22*(6), 329-340. doi:10.1007/s00787-013-0375-0

Filipek, P. A., Accardo, P. J., Ashwal, S., Baranek, G. T., Cook, E. H., Jr., Dawson, G., . . . Volkmar, F. R. (2000). Practice parameter: screening and diagnosis of autism: report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology, 55*(4), 468-479.

Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *J Child Psychol Psychiatry, 54*(5), 582-590. doi:10.1111/jcpp.12008

Hayes, J., Ford, T., Rafeeque, H., & Russell, G. (2018). Clinical practice guidelines for diagnosis of autism spectrum disorder in adults and children in the UK: a narrative review. *BMC Psychiatry, 18*(1), 222. doi:10.1186/s12888-018-1800-1

Insel, T. R. (2014). The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *Am J Psychiatry, 171*(4), 395-397. doi:10.1176/appi.ajp.2014.14020138

Insel, T. R., & Cuthbert, B. N. (2015). Medicine. Brain disorders? Precisely. *Science, 348*(6234), 499-500. doi:10.1126/science.aab2358

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry, 17*(12), 1174-1179. doi:10.1038/mp.2012.105

Lavelle, T. A., Weinstein, M. C., Newhouse, J. P., Munir, K., Kuhlthau, K. A., & Prosser, L. A. (2014). Economic burden of childhood autism spectrum disorders. *Pediatrics, 133*(3), e520-529. doi:10.1542/peds.2013-0763

Lord, C. (1991). Methods and measures of behavior in the diagnosis of autism and related disorders.

Psychiatr Clin North Am, 14(1), 69-80.

Lord, C., & Jones, R. M. (2012). Annual research review: re-thinking the classification of autism spectrum disorders. *J Child Psychol Psychiatry, 53*(5), 490-509. doi:10.1111/j.1469-7610.2012.02547.x

Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D. M., . . . Risi, S. (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Arch Gen Psychiatry, 69*(3), 306-313. doi:10.1001/archgenpsychiatry.2011.148

Mahjouri, S., & Lord, C. E. (2012). What the DSM-5 portends for research, diagnosis, and treatment of autism spectrum disorders. *Curr Psychiatry Rep, 14*(6), 739-747. doi:10.1007/s11920-012-0327-2

Ozonoff, S. (2012). Editorial perspective: Autism spectrum disorders in DSM-5--an historical perspective and the need for change. *J Child Psychol Psychiatry, 53*(10), 1092-1094. doi:10.1111/j.1469-7610.2012.02614.x

Tanguay, P. E. (2011). Autism in DSM-5. *Am J Psychiatry, 168*(11), 1142-1144. doi:10.1176/appi.ajp.2011.11071024

Towle, P. O., & Patrick, P. A. (2016). Autism Spectrum Disorder Screening Instruments for Very Young Children: A Systematic Review. *Autism Res Treat, 2016*, 4624829. doi:10.1155/2016/4624829

Vohra, R., Madhavan, S., Sambamoorthi, U., & St Peter, C. (2014). Access to services, quality of care, and family impact for children with autism, other developmental disabilities, and other mental health conditions. *Autism, 18*(7), 815-826. doi:10.1177/1362361313512902

Walter, H. (2013). The third wave of biological psychiatry. *Front Psychol, 4*, 582. doi:10.3389/fpsyg.2013.00582

Zuckerman, K. E., Mattox, K., Donelan, K., Batbayar, O., Baghaee, A., & Bethell, C. (2013). Pediatrician identification of Latino children at risk for autism spectrum disorder. *Pediatrics, 132*(3), 445-453. doi:10.1542/peds.2013-0383