FACIAL EXPRESSION PROCESSING IN AUTISM SPECTRUM DISORDER AS A FUNCTION OF ALEXITHYMIA: AN EYE MOVEMENT STUDY

by

Brian Escobar

A Dissertation Submitted to the Faculty of

Charles E. Schmidt College of Science

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Florida Atlantic University

Boca Raton, FL

December 2023

FACIAL EXPRESSION PROCESSING IN AUTISM SPECTRUM DISORDER AS A FUNCTION OF ALEXITHYMIA: AN EYE MOVEMENT STUDY

by

Brian E Escobar

This dissertation was prepared under the direction of the candidate's dissertation advisor, Sang Wook Hong, Department of Psychology, and has been approved by all members of the supervisory committee. It was submitted to the faculty of the Charles E. Schmidt College of Science and was accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

SUPERVISORY COMMITTEE:

Sang Wook Hong
Sang Wook Hong (Dec 12, 2023 10:15 EST)

Sang Wook Hong, Ph.D.
Dissertation Advisor

Gizelle Anzures
Gizelle Anzures (Dec 12, 2023 22:09 EST)

Gizelle Anzures, Ph.D.

Chad E. Forbes, Ph.D.

Nancy Aaron Jones (Dec 13, 2023 10:59 EST)

Nancy Aaron Jones, Ph.D.

Lira Yoon

Lira Yoon, Ph.D.

Alan W. Kersten Ph.D.
Chair, Department of Psychology

Valery E. Forbes, Ph.D.
Dean, Charles E. Schmidt College of Science

Robert W. Stackman Jr., Ph.D.
Dean, Graduate College

December 13, 2023

Date

ACKNOWLEDGEMENTS

I want to offer my gratitude to Dr. Sang Wook Hong for his dedication to helping me complete this study and for affording me the opportunity to work in his lab. Additionally, I would like to express my appreciation towards Dr. Summer Sheremata and Dr. Gizelle Anzures for their kind words during my thesis defense, which has helped bolster my confidence as an academic. Thank you to my committee for helping me get this far. Karine Grace, Brittany Menoutis, Lindy Polikar, and John Whu, thank you all for helping me collect the data reflected in this dissertation. Finally, I am thankful to Florida Atlantic University for providing the necessary resources to conduct research.

ABSTRACT

Author: Brian E. Escobar

Title: Facial Expression Processing in Individuals with Autism Spectrum Disorder as a Function of Alexithymia: An Eye Movement Study

Institution: Florida Atlantic University

Dissertation Advisor: Sang Wook Hong, Ph.D.

Degree: Doctor of Philosophy

Year: 2023

The perception and interpretation of faces provides individuals with a wealth of knowledge that enables them to navigate their social environments more successfully. Prior research has hypothesized that the decreased facial expression recognition (FER) abilities observed in autism spectrum disorder (ASD) may be better explained by comorbid alexithymia, the alexithymia hypothesis. The present study sought to further examine the alexithymia hypothesis by collecting data from 59 participants and examining FER performance and eye movement patterns for ASD and neurotypical (NT) individuals while controlling for alexithymia severity. Eye movement-related differences and similarities were examined via eye tracking in conjunction with statistical and machine-learning-based pattern classification analysis. In multiple different classifying conditions, where the classifier was fed 1,718 scanpath images (either at spatial, spatial-temporal, or spatial-temporal-ordinal levels) for high-alexithymic ASD, high-alexithymic

NT, low-alexithymic ASD, and low-alexithymic NT, we could accurately decode significantly above chance level. Additionally, in the cross-decoding analysis where the classifier was fed 1,718 scanpath images for high- and low-alexithymic ASD individuals and tested on high- and low-alexithymic NT individuals, results showed that classification accuracy was significantly above chance level when using spatial images of eye movement patterns. Regarding FER performance results, we found that ASD and NT groups performed similarly, but at lower intensities of expressions, ASD individuals performed significantly worse than NT individuals. Together, these findings suggest that there may be eye-movement-related differences between ASD and NT individuals, which may interact with alexithymia traits.

DEDICATION

This dissertation is dedicated to my family, particularly my younger brother, Aidan, who inspired me to research autism. I'd also like to dedicate this work to The Campout Crew for staying in close contact and remaining close friends since middle school. Additionally, someone please donate a kidney to Josh Covey (DOB: 07/19/1991). A direct match would be blood type O. All blood types could help through the Paired Exchange Program. To be tested, call UPMC Pinnacle Harrisburg (877-778-6110). Finally, this dissertation is dedicated to baby Hazel, the Philadelphia Eagles, and Kyle Busch for being my sources of light and hope in an otherwise dark and defeating academic world.

FACIAL EXPRESSION PROCESSING IN INDIVIDUALS WITH AUTISM SPECTRUM DISORDER AND ALEXITHYMIA

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. SIGNIFICANCE OF FACE PROCESSING

An abundance of information can be garnered when an individual views a face. When individuals look at a face, they must deduce two key pieces of information. One key piece of information is who the face belongs to (identity), and the second is the emotional information which the face is trying to convey or express. It has been argued that highly specialized systems have likely evolved to process facial information because facial identity and expression processing, recognition, and categorization can occur at such rapid paces (Posamentier & Abdi, 2003). The use of this information enables us to better navigate our social environments and increase our odds of survival. For example, recognizing that another individual is angry by recognizing pursed lips, a furrowed inner brow, and relaxed lower eyelids can enable us to disengage socially from an individual who is expressing emotions of anger and may potentially be a threat. Regarding facial expression information, it should also be noted that expressions likely have high degrees of evolutionary and social importance, such that individuals across various cultures can recognize numerous emotions universally. These expressions have been dubbed basic expressions and are comprised of the emotional expressions of anger, disgust, fear, happy, sad, and surprise (Ekman & Friesen, 1971). Darwin (1872) was the first to propose that emotional facial expressions may possibly hold relevant evolutionary information, and the recognition of certain expressions may not vary across different cultures. These ideas were further substantiated by work conducted by Ekman and Friesen (1971) that examined if individuals belonging to preliterate cultures who had minimal interactions with literate Western and Eastern cultures could associate emotional expressions with narratives that conveyed emotional information. In the task, the individuals of the preliterate culture were told a short story that contained emotional information and then were tasked with choosing which of three emotional expression faces would best belong to the emotional short story (e.g., correct response for a short story that conveyed emotional information of anger would be denoted by choosing the emotional expression face that portrayed anger. An incorrect response would be choosing either of the other two emotional expression faces, such as happy or sad). Results demonstrated that individuals who belonged to the preliterate culture could choose the correct emotional expression face for the corresponding emotional short story significantly higher than chance. These results would suggest that facial expressions likely hold evolutionary and social significance such that there are expressions that may be preserved across numerous cultures despite any potential influences from other outside cultures. A cadaver study conducted by Waller et al. (2008) has shown support for the argument that there are likely universals regarding emotional facial expressions and provides an argument for how these universal basic expressions are generated. In this study, the facial muscles of 18 cadavers were examined and the results showed that although there could be asymmetries in the facial muscles across the cadavers, there was less asymmetry regarding facial muscles that are implicated in the generation of facial expressions (e.g., the zygomaticus major used to pull up the corner of the lips in an expression of happiness). These results would likely suggest that there is selection pressure for the preservation of facial muscles used for facial expression due to their potential evolutionary and social importance.

Work has also been conducted to provide us with insight into where in the brain the distinct coding of facial identity and expression may take place. In the case of individuals with prosopagnosia, these individuals are unable to recognize and categorize faces based on identity yet may still be able to recognize the expressions presented by another's face (Posamentier & Abdi, 2003). This would suggest that facial identity and expressions may possibly be encoded by distinct and specialized systems within the brain. Regarding the impairment of recognizing facial identities, there seem to be two forms of prosopagnosia, which are caused by damage to different regions of the brain. Work by Farah (1991) suggested that one form of prosopagnosia, known as "associative prosopagnosia," is typically caused by bilateral damage to the inferior occipital and temporal visual cortices and to areas of the posterior temporal region. This form of agnosia causes the patients to have perceptual systems that appear to be healthy enough to allow for adequate perception to occur, yet they still fail to recognize facial identities. Damasio et al. (1990) noted that the other form of prosopagnosia, "appreciative prosopagnosia" is marked by damage to the right visual association cortices in the occipital and parietal brain regions. Damage to these areas causes inadequate performance in visual perception; thus, the patient cannot see the face in a typical manner and is unable to recognize the identity of the face. Although we understand that facial identity may be given its own unique systems or subsystems within the brain for encoding, work has also been done to provide us insight into what regions of the brain are implicated in the encoding and processing of facial expressions.

Hasselmo et al. (1989) examined the neuronal responses to identity and expression in the temporal visual cortex of macaque monkeys. The monkeys viewed images of three monkeys that displayed three different expressions, and 45 neurons were selected for recording. Results showed that of the 45 neurons examined, 15 responded to identity independently of expression, and nine responded to expression independently of identity. These results regarding identity-selective and expression-selective neuronal responses were then replicated in humans by a patient study conducted by Ojemann et al. (1992). They examined neuronal responses of 21 different populations of neurons across the right superior and middle temporal gyri of 11 patients undergoing craniotomy. Participants of the study were asked to complete two tasks, one on expression labeling and the other on identity matching. When completing both tasks there were significant neuronal activity changes, 62% change during identity matching and 52% change during expression labeling. The results provided support for the findings of Hasselmo et al. (1989) by again showing differences in neuronal-selective responses across expression and identity. Additionally, it should be noted that in the middle temporal gyrus, there were more localized patterns of response during the expression task. This may suggest specificity and differences in neuronal responses surrounding faces regarding identity and expression.

The following chapters will detail various aspects of face processing, such as proposed models of face perception, the coding of identity and emotional/expression information, the behavioral and neural responses related to the temporal dynamics and accuracy of facial expression classification, techniques used when examining face processing, and disruptions of face processing in individuals with autism spectrum disorder (ASD) and alexithymia. Followed by an experimental proposal that uses behavioral, eye-tracking, and computational methods to examine potential differences in face processing amongst ASD, alexithymic, and neurotypical (NT) individuals.

CHAPTER 2. MODELS AND FRAMEWORKS OF FACE PERCEPTION

Numerous models have been suggested regarding the potential separation of how aspects of faces may be processed and interpreted, such as those relating to facial identity and facial expression. Models like the ones proposed by Bruce and Young (1986) and Haxby et al. (2000) share commonalities in their ideas of how facial identity and expression are likely processed separately through different cognitive mechanisms or pathways. However, where they differ is that Haxby and colleagues provide neurotopical support for this claim, while the Bruce and Young model does not. There are, however, frameworks that critique these long-applied models, such as the one from Duchaine and Yovel (2015), that provides new interpretations of how facial identity and expression information may be encoded, interpreted, and applied from distinct neural pathways.

2.1. Bruce and Young Model

Bruce and Young (1986) aimed to create a model of face perception that pulled together and extended information from previous models (e.g., Bruce, 1979, 1983; Rhodes, 1985) while also organizing a framework for how face recognition occurs in the context of familiar vs. unfamiliar faces and other aspects of faces processing. Within this model, they suggest that the information obtained through the structural encoding of faces allows for the analysis of facial speech, expression, and recognition units. And that recognition is achieved when previously-stored recognition units match the information gathered from the structural encoding of a face. Additionally, they suggest there are other distinct categories of information or codes that are products of face processing. The codes they establish are pictorial, structural, visually derived semantic, identity-specific semantic, name, expression, and facial speech codes.

2.1.1. Face Codes of Bruce and Young Model

Pictorial codes are only obtained through the viewing of face pictures and are representative of the description of the picture. Thus, pictorial codes do not hold information that is necessarily face-specific. Rather, these codes hold information about the lighting of the picture, the color of the picture, as well as the pose and expression portrayed within the picture. As previously mentioned, structural codes lend to the recognition of familiar faces. Structural codes provide information not just on specific face areas, such as the eyes and mouth, etc., but also on the configuration of these features as well. For example, being able to recognize the identity of a familiar face when it has been caricatured. That is, the features themselves have been distorted, but the configuration remains the same, yet accurate recognition is still able to occur.

Another influence on structural code information is the familiarity of a face. For example, when an individual is more familiar with a face, they are more likely to obtain structural code information from internal facial features, such as the eyes, nose, and mouth. However, when an individual is less familiar with a face, there is an equal emphasis on obtaining structural code information via both internal and external features of the face. That is, the individual is just as likely to gather information that is more related to changeable features of recognition, such as the hair of another individual. One way in which the remembering of a familiar face can be facilitated is by processing, recognizing, and interpreting aspects of the face, such as sex, age, honesty, intelligence, etc. These aspects of face perception are known as visually derived semantic codes.

If visually derived semantic codes can facilitate the remembering of an unfamiliar face, then familiar face descriptions are likely obtained through identity-specific codes. Identity-specific codes are related to aspects such as the occupation of a familiar individual, the friends of the familiar individual, what clothes they commonly wear, etc. However, it should be mentioned that there have been arguments made against such discrete types of semantic information-based codes used for recognition. Rhodes (1985) has argued that face-specific semantic code information falls on a continuum that is related to meaningfulness. For example, a highly encountered familiar face that is important to an individual is likely to hold more information pertaining to meaningfulness than a face that has only been encountered once. Bruce and Young (1986) argue against such notions by stating that these distinct semantic codes allow you to accurately describe the occupation of a familiar face and that one would be unable to do so if the face belongs to an unfamiliar face. Thus, only through being familiar with a face are you able to recall semantic information that is related to its identity, hence why there are likely discrete categories of semantic-based codes.

Identity-specific code information should not be confused with items such as the name of an individual. Rather, Bruce and Young (1986) argue that there are distinct name codes related to face perception that are concerned with this information. Identity-specific and name codes differ because it is possible to look at a face and accurately remember their occupation (an identity-specific code) but be unable to accurately recall what their name is (name code). The last two face codes proposed are argued to not be important in terms of face recognition. These are expression codes and facial speech codes. Expression codes are obtained when viewing the shapes of facial features of another as they intend to portray an emotional state. Facial speech codes are obtained via the viewing of lip and tongue movements. Within their model, they argue that identity and expression recognition occur within separate functional routes, but they direct their focus mainly on the recognition of identity.

2.1.2. Functional Components of Bruce and Young Model

In the Bruce and Young (1986) model, face recognition units send information to the cognitive system. Within these face recognition units is information that pertains to the structural codes. The strength of this signal from a face recognition unit to the cognitive system can be modulated by the degree of resemblance between the input obtained via structural encoding and the already stored information. Additionally, greater responses can be obtained through priming, recency, and anticipation of seeing a face. Face recognition units have access to person identity nodes, which are identity-specific semantic codes held within associative memory. Bruce and Young (1986) suggest that there is one person identity node for each person an individual knows and that each node holds identity-specific semantic codes.

It should be noted that a face recognition unit only activates when any angle of a person's face is viewed, and that identity-specific semantic code activation can occur when any semantic information pertaining to an individual is perceived, such as face, voice, hair, clothing, etc. Therefore, the overall distinction is that identity-specific semantic code activation is related to the moment in which person recognition occurs, rather than face recognition. All episodic and associative information that is not kept within the person's identity nodes is included in the cognitive system of the Bruce and Young model. They also suggest that this system plays a role in the direction of attention to other system components, such as directed visual processing, and analyses of expression and facial speech. However, as previously mentioned, this model does not offer much in the explanation of perception, recognition, and interpretation of these latter-mentioned facial aspects since it is mainly focused on how facial recognition pertaining to identity is achieved.

2.1.3. Face, Object, and Word Recognition in The Bruce and Young Model

Within their model, Bruce and Young (1986) also compare similarities and differences between face, object, and word recognition. They highlight that the routes involved in the recognition of these categories are highly similar; however, the main differences between these are highlighted when examining the speed at which an individual naming response or a semantic categorization decision response can be made between the categories. Each of these categories begins with an input code (a form-specific code for an object, an orthographic code for a word, and a structural code for a face), which then feeds into specific recognition units for each category (the recognition units for objects are known as pictogens, words are logogens, and faces are face recognition units). At this point, responses for these categories can differ. For words, they are already able to be named, but for objects and faces, this cannot occur yet. The next step in this route for recognition involves identity-specific semantic representations. It is here that all categories can be semantically categorized, but for faces and objects to be named, they must first gather information from name codes.

2.2. Haxby and Colleagues' Model of Face Perception

Haxby et al. (2000) proposed a model of face perception, where instead of having a model that is suggested to be specialized for face perception, it is instead rather mediated by distributed processing. Their model minimizes limitations of the Bruce and Young (1986) model by also providing neuroimaging evidence to suggest that distinct neural representations are what mediate different aspects of face perception. Their model is mainly concerned with variant and invariant aspects of face perception and identification. The invariant aspects of faces are ones that leave facial structure intact during the movement of social communications, such as during expression, or changes in eye and mouth movement. These areas of the face that move to facilitate social communication potentially further are the variant or changeable aspects of a face. They argue that if the face perception system represents both invariant and variant facial aspects, then representations of identity are independent of variant facial aspect representation. That is, identity must be independent from more socially communicative based aspects of faces, such as expression, eye gaze, and speech-related movements. However, it should be made clear that Haxby et al. (2000) address that within their model, it remains unclear how separate the functional roles are for each of the regions implicated in this model.

2.2.1. Core System of Face Perception – Haxby and Colleagues

Haxby et al. (2000) propose that within the occipitotemporal visual extrastriate cortex are three bilateral regions that make the core system for face perception. These three regions are within the inferior occipital gyri, lateral fusiform gyrus, and superior temporal sulcus (STS). They argue that the lateral fusiform gyrus may be involved with identity-based representations, that STS is more involved in the representation surrounding changeable aspects of faces, and that inferior occipital gyri likely provides inputs to both areas because of its anatomical location.

Haxby et al. (2000) provide evidence from functional brain imaging studies that face perception evokes responses from the lateral fusiform gyrus and that although this activity is bilateral, it is more regularly found within the right hemisphere (e.g., Kanwisher et al., 1997). Additionally, it has been demonstrated that these evoked responses are greater for faces than for non-face object stimuli (e.g., Kanwisher et al., 1997). This preference for face vs. non-face object stimuli has led to the idea that this area is specialized for face perception and has been coined, the fusiform face area (FFA). However, Haxby et al. (2000) pointed out that there is work that demonstrates that when individuals are tasked with attending to variant aspects of the face, such as the direction of eye gaze, there is a reduction in the FFA's magnitude of response (Hoffman & Haxby, 2000). This would suggest that this area does not play a role in all aspects of face perception and instead is more heavily implicated in the perception of invariant aspects of faces. Haxby et al. (2000) mention that other neuroimaging work, as well as electrophysiological work, has suggested that areas of the core system of face perception have also been shown to be face-responsive. These areas are the lateral inferior occipital gyri and the posterior superior temporal sulcus (pSTS) (Gorno-Tempini et al., 1998; Haxby et al., 1999; Hoffman & Haxby, 2000; Leveroni et al., 2000).

2.2.2. Extended System of Face Perception – Haxby and Colleagues

The proposed extended system of face perception from Haxby et al. (2000) suggests that there are brain regions that are involved in other aspects of cognition but work with the core system of face perception to aid in the significance of information garnered from looking at a face. Within these other aspects of cognition, stimuli recognition is influenced by semantic information and information stemming from the senses. An example is how semantic context and perceived lip movements influence the perception of speech, which can be further highlighted by the McGurk effect (McGurk & Macdonald, 1976). Regarding face perception, recognition of expression can be influenced by the emotional tone of the expression.

Expression research can be used to demonstrate how there may be an extended system of face perception. Magnetoencephalography (MEG) work done by Streit et al. (1999) has demonstrated that in the pSTS and later in the right amygdala, there were stronger activations of these areas when performing an emotional judgment task vs. a face detection task. It has been suggested that outside of the perception of fear, the amygdala is also implicated in the processing of facial information, that is critical for social cognition. Work by Baron-Cohen (1995) has demonstrated that the perception of the eye region when judging the state of mind of others can lead to an increase in amygdala activity. Within this study, ASD participants had less amygdala and inferior frontal cortex activity but greater activation of the superior temporal region. These results may suggest that atypical activations within these areas can lead to impaired social cognition.

Adolphs (1999) demonstrates another brain region of the proposed extended system that may be implicated in face processing, the somatosensory cortex. Adolphs (1999) argues that the complex emotion recognition for facial expressions is likely facilitated through the incorporation of the somatosensory cortex. That is, when tasked with emotional face judgments, performance is likely dependent upon two strategies. One strategy that could be deployed is potentially making judgments based upon knowledge surrounding known facial configurations for emotional expression (e.g., a down-turned mouth likely is a signal for sadness). Another strategy may be to create somatosensory images about how one may feel if they were making the facial expression displayed to them. This strategy is potentially useful when not having any prior knowledge about why the face may be displaying the expression presented. Adolphs (1999) argues that this second strategy can be implemented either covertly or overtly. This argument that the somatosensory cortex may be part of the extended system of processing comes about through Adolphs' earlier patient studies. Here, it was found that emotion recognition relied critically on the integrity of somatosensory-related cortices in the right hemisphere, particularly when tasked with judging complex blends of multiple emotions (Adolphs, 1998; Adolphs, 1999).

2.3. Critiques of Haxby and Colleagues' Model

As previously mentioned, the model of face perception proposed by Haxby et al. (2000) was created to minimize the limitations of Bruce and Young's (1986) model. Their work provided results from neuroimaging studies that would suggest that distinct aspects of face perception are mediated by the distributed processing of faces. However, other studies have suggested that the Haxby et al. (2000) model may also need to undergo modifications to address its own limitations. For example, it may be that face perception does in fact rely on two distinct neural pathways. A review by Duchaine and Yovel (2015) has suggested that a ventral stream for face processing may be used for form representation information and that a dorsal stream may be implemented in the processing of motion and form face-related information.

Haxby and Colleagues' (2000) model of face perception proposed that due to the anatomical location of the occipital gyri, the Occipital Face Area (OFA) likely acts as a key area of face information input for a suggested distributed face-processing network. However, patient studies on those with damage to brain areas, such as the OFA and FFA, have suggested that there may be multiple facial information input areas within a face-processing network. For example, an fMRI study from Sorger et al. (2007) examined for the first time the functional neuroanatomy of an acquired prosopagnosic, patient PS. Within their study, patient PS and two controls were subjected to a one-back matching task, in which they were tasked with responding to any time a face or object was presented two times in a row. Results of fMRI data demonstrated that lesions to patient PS's OFA did not disrupt the preferential processing of face-related information in areas such as the FFA and pSTS when compared to controls. Further evidence of face-related information still being able to reach face-specific areas within a face-processing network has been shown in work, as seen in a study conducted by Yang et al. (2016). In this study, they sought to examine the nature of identity representations within the anterior temporal lobe (ATL) and its relationship to posterior regions of face perception, such as the OFA and FFA. They used patient Galen within this study due to lesioning of their right OFA and right FFA. Despite lesioning to these posterior brain areas implicated in face perception, face-related information was still able to reach the face-selective region of the anterior temporal lobe in patient Galen. Together, results from both studies would suggest that the argument presented by Haxby et al. (2000), where face-related information is likely distributed to other brain areas through the OFA, may not hold up and that there may in fact be multiple pathways for face-related information input into a face-processing model.

Finally, in a review from Duchaine and Yovel (2015), they propose a revised neural framework for face processing. Within this framework, they suggest that there may be two distinct streams within a face-processing network. One is a dorsal stream compromised of the posterior superior temporal sulcus face area (pSTS-FA), the anterior superior temporal sulcus face area (aSTS-FA), and the inferior frontal gyrus face area (IFG-FA). Areas within this stream have been shown to display stronger responses to dynamic or variant representations of faces, such as eye gaze and expression. They suggest that this stream would likely play a role in social interactions, which are dependent upon the varying aspects of face perception, such as mouth and eye movements. Whereas the ventral stream has been shown to represent invariant aspects of face perception, such as identity, sex, and age judgments, and is comprised of the OFA, FFA, and anterior temporal lobe face area (ATL-FA). They argue that in this stream, the OFA and FFA are responsible for the surface and structural representations of face processing, while in the next area in the stream, the ATL-FA is responsible for matching these representations to previously experienced faces to aid in the proper identification of the face. Certainly, further investigations should be performed as to the extent to which this newer model can explain multiple aspects of face perception.

CHAPTER 3. CODING OF FACIAL IDENTITY AND EXPRESSION

3.1. Norm-Based Coding

Theories of face processing propose that the faces are processed and coded in a prototype-referenced, multidimensional framework. This type of "norm-based" coding allows for the proper identification of facial identities across individuals by comparing the variations and deviations of the presented face from an average prototype face within the individual's consciousness, also known as a face space (Gwinn et al., 2018; Rhodes & Jeffery, 2006). Within the face space of norm-based or two-opponent pool coding, each value along a facial feature dimension (e.g., eye height) is coded by the relative input of the face compared to a norm. But in the two-pool model, this relative input is coded by the output of two pools of neurons that are oppositely tuned. One of these pools will respond more strongly to higher value ranges along a dimension (e.g., eyes higher up on the face) and less strongly to lower value ranges along a dimension (e.g., eyes being lower on the face). The other pool within this system will respond in an opposite fashion, where it will respond minimally to higher values and more strongly to lower-end values. The face is thus processed, perceived, and identified when the pools process values above and/or below the average value dimensions of the face space.

It has also been postulated that norm-based coding may occur during the perception, proper identification, and recognition of facial expressions/emotions (Gwinn et al., 2018; Skinner & Benton, 2010). One of the aims of the study conducted by Gwinn et al. (2018) was to examine if there were asymmetrical neural responses for facial expressions (i.e., happy, angry, and fearful faces were used for the purpose of the study) and neutral expressions. Within one of the conditions of the experiment, participants were tasked with viewing photo-realistic images under the presentation method of Fast Periodic Visual Stimulation (FPVS). The images alternated between those displaying expressions and those displaying neutral faces. Each trial had the images displayed six times per second (i.e., held constant at 6 Hz) for 20 seconds, with the alternations of the two image sets causing repetitions of the same faces at 3 Hz. In both conditions, significant neural responses from the right occipito-temporal region were observed at 3Hz, recorded by the electroencephalogram (EEG) system. Thus, confirming that the expressions and neutral faces produced asymmetric responses. These asymmetric responses when processing expressions vs. a “norm” face (i.e., the neutral face) would imply possible evidence for the norm-based coding of expressions. The rationale being that if norm-based coding is implicated to be occurring, we would expect to see different neuronal responses when viewing an expression that was not the prototype face because of the activation of different neuronal pools. We would expect to observe larger response patterns when viewing an expression that has values either above or below the average value of the prototype-referenced neutral face, activating these different neuronal pools. The viewing of the neutral face should not cause such large neuronal responses because it has been postulated that the neutral face holds values close to the average values of the prototype face. Results of their experiment were consistent with this rationale.

It should also be noted that adaptation aftereffects may provide support for the argument of norm-based coding occurring during the processing of faces. When individuals view a visual stimuli or stimulus for a prolonged period, a reduction in sensitivity to what is being viewed occurs. This reduction in sensitivity can cause a shift in the perception of stimuli that are viewed immediately after adaptation to the original stimuli takes place. For example, if individuals are asked to view a video of a waterfall flowing downwards for a prolonged period and are then presented with a static picture of the waterfall, their perceptual experiences may bias them to perceive the illusion of a waterfall that has now begun to flow upwards instead. These adaptation aftereffects have been shown to occur in a variety of different low-level visual features, such as size (Blakemore & Sutton, 1969), color (McCollough, 1965; Tregillus & Engel, 2019), in complex visual stimuli like faces (Benton et al., 2007; Hong & Yoon, 2018; Leopold et al., 2001; Lithfous & Rossion, 2018; Skinner & Benton, 2010).

Regarding adaptation aftereffects concerning faces, this process is known as the Face Adaptation Aftereffects (FAA). The FAA may provide us with revelations into how faces are processed and encoded. A study conducted by Webster et al. (2004) examined the adaptation aftereffects of faces and their influences on judgments of natural face categories, such as gender, ethnicity, and expression. Stimuli were generated in a continuum manner by morphing pairs of each of the aforementioned natural face categories. At the center of the continuum was an ambiguous stimulus of the two morphed images, and on the extreme ends of the continuum were the original images used to create the morph. Results showed that after adaptation to a male face, the ambiguous image was perceived as female and that adapting to the female face caused the ambiguous image to appear more male. Similar results were found for the other two face categories as well. Adapting to a Japanese or Caucasian face was shown to significantly bias the perception of the neutral image. Meaning that adapting to a Japanese face caused the neutral image to appear more Caucasian and that adapting to a Caucasian face caused the neutral image to appear more Japanese. These aftereffects were also found concerning expression, such that adaptation in an angry-happy morph caused adaptation to the angry face to make the neutral image appear happier, and adaptation to the happy face to make the neutral image appear angrier. These results suggest that adaptation may cause a re-centering of the face space relative to the adapting facial configuration and provide possible evidence for the implied importance of coding of various aspects of faces.

A study conducted by Skinner and Benton (2010) also examined FAA and the representation of facial expressions. Results of this study would go on to present further insights into how the FAA may provide us with revelations of how faces are processed and encoded, and how norm-based coding is involved in such areas of attention and perception. Participants within the study were adapted to the anti-expression counterparts of the six basic expressions at both 50% and 100% strengths, then their perceptions of expressions were recorded after being presented with a prototype neutral face. Anti-faces and the aforementioned emotional anti-faces/anti-expressions are generated by taking a facial identity (e.g., Kyle) or expression (e.g., happy) and running them through a face space, through and across a said neutral face. The resulting stimulus is a face created with the opposite facial shape of the original face (i.e., anti-Kyle or anti-happy) (Gwinn et al., 2018). Results of Skinner and Benton (2010) showed a significant effect of anti-expression adaptation on basic expression perception, meaning that adapting to anti-expression caused a greater perception of its expression counterpart when viewing the prototype neutral face (e.g., adaptation to anti-happy caused the participant to perceive the prototype as happy). Additionally, when examining the effect of adapter strength on aftereffect strength, it was found that aftereffect strength was greatest when adapting to the 100% anti-expression vs. the 50% expression. These results would suggest that facial processing occurs within a prototype-referenced multidimensional framework. This is because, within a norm-based coding model of face perception, one would expect that aftereffect strength would increase as the adapter stimulus became further and further away from an average prototype face. The findings that stronger adapters lead to a greater aftereffect and biased the perception of anti-expressions are representative of a prototype-referenced framework, thus providing support that facial expressions are coded in a norm-based manner.

3.2. Exemplar-Based Coding of Identity and Expression

Aside, from norm-based coding, another form of face processing has been posited. This other form of processing is known as exemplar-based coding. Rather than comparing variations and deviations from an average proto-type face as seen in norm-based coding, exemplar-based coding posits that individuals can create an average proto-type face, but it holds no value in discerning variations among facial identities or emotional expressions. Instead, regarding facial identity, faces are processed and coded in relation to their absolute values on each dimension of a proposed face space (Jeffery et al., 2011). Thus, rather than coding taking place relative to the prototype face, faces are encoded based upon their exemplars from other faces that were processed and perceived in the past. Regarding exemplar-based coding, values along the trajectory of the face space are coded by various pools of neurons. Each of the neuronal pools has bell-shaped tuning centered over a different value. Similar values (e.g., similar mouth height) in this model will generate responses from the same set of neurons, but values that are different will generate responses in non-overlapping sets of neurons. The face is thus processed, perceived, and identified when pools process values above and/or below the absolute value dimensions of the face space.

To examine if exemplar-based coding is implicated in the coding of identity, Robbins et al. (2007) utilized the dimension of eye height in symmetric (i.e., both eyes up or down) and asymmetric (i.e., one eye up and the other one down) distortions at various intensities of the distortions (i.e., lower-value intensities are those where eye-height is adjusted minimally, and higher-value intensities are those where eye-height is adjusted more extremely), and their influences on perceptions of average faces as well as learned identities. They predicted that in an exemplar-based coding model, stronger aftereffects should emerge when the adaptor is at a lower intensity than at a higher one due to the lower-intensity distortion lying closer to the average face. This is because the lower-intensity distorted image should activate the same neuronal pool as the average face if they carry similar absolute values within a face space. They also made predictions that a norm-based coding model would portray the opposite results, where higher-intensity distorted adaptors should produce a stronger aftereffect because the neurons would be tuned to more extreme values. Their results showed support for the norm-based coding of faces, where more extreme distortion intensities lead to greater adaptation aftereffect sizes for the aforementioned facial attributes.

To date, numerous studies provide similar evidence against an exemplar-based coding model concerning both facial identity and expression. Regarding the processing of facial identity and FAA, work has suggested that two opponent (norm-based) coding may be more heavily implemented than multichannel (exemplar-based) coding (Jeffery et al., 2010; Jeffery et al., 2011; Leopold et al., 2006; Robbins et al., 2007). Additionally, evidence suggests that norm-based coding may also be implicated in the processing of expressions as well (Burton et al., 2013; Neth & Martinez, 2009). However, a study by Ross et al. (2014) provides computational evidence of exemplar-based and two-opponent pool coding models being implicated in facial identity processing, while showing less support for the traditional norm-based coding model.

Ross et al. (2014) argued that the general trend of research supporting a norm-based coding model that also argues against an exemplar-based model is missing supporting evidence from mathematical and computational models for facial identity aftereffects. To minimize this limitation, they implemented simplified versions of norm- and exemplar-based coding models and tested their predictions of facial identity aftereffects. Each computational model was used to examine if they could obtain predictions of the observed pattern of face identity aftereffects across multiple parameters. Key parameters they were interested in examining were the number of dimensions in the perceptual input and face-space representation layers, the broadness of tuning curves, and the strength of adaptation. These parameters were adjusted incrementally to investigate their influence on model predictions. Additional parameters, such as response mapping across all models and adaptation scaling for the traditional norm-based model, were adjusted to generate correspondence between model predictions and previously recorded behavioral data. These computational models were applied to three different behavioral-study paradigms, the anti-face adaptation paradigm seen in Leopold et al. (2001) to examine anti-face adaptation, an extension of the Leopold et al. (2001) anti-face adaptation paradigm from Leopold and Bondar (2005) to examine the effect of adaptor position relative to the average face, and a modified anti-face adaption paradigm from Rhodes and Jeffery (2006) to examine adaptation to opposite and non-opposite morphs of facial identities.

In the anti-face adaptation paradigm of Leopold et al. (2001), the goal was to investigate if adapting to an anti-face would facilitate recognition of the corresponding target face (e.g., would adapting to anti-Adam facilitate recognition of Adam?). Recognition of identity was investigated in adapting and non-adapting conditions along a morph trajectory of faces that reached from the anti-face to the target face. Results of the simulation from Ross et al. (2014) were consistent with the behavioral results of Leopold et al. (2001), where adapting to a matching anti-face facilitated identity recognition and adapting to a non-matching anti-face impaired identity recognition. These results were consistent across the exemplar-based, traditional norm-based, and two-opponent pool models, and multiple parameter values for each model. These results were interpreted as the behavioral results of Leopold et al. (2001) could not accurately differentiate between norm-based and exemplar-based coding models due to both models making similar predictions surrounding anti-face adaptation.

In their next simulation, Ross et al. (2014) examined the effect of adaptor position on facial identity adaptation. In the study by Leopold and Bondar (2005) adaptation was compared in moderate anti-face strength conditions and average face conditions while being compared to a baseline of no adaptation. Results of this study provided further support for the findings of Leopold et al. (2001), where adaptation to the moderate strength anti-face resulted in a strong biasing of identifying the target face more correctly compared to baseline conditions. In addition, adapting to the average face caused non-significant biasing of correct identifications. It has been argued that these results are supportive of norm-based coding because if biasing only occurs in an opposite trajectory manner, then a face norm must be held in the face space because an opposite would only be possible if there is a norm to base an opposite around. Computational results of Ross et al. (2014) provided further support for biasing occurring in an opposite trajectory fashion, as seen by larger adapter effects being produced when adapting to moderate strength anti-faces, and little if any adapter effects when adapting to average faces. However, it should be noted that this biasing pattern was seen across all three models, meaning both the norm-based models and the exemplar-based model.

In their last set of simulations, Ross et al. (2014) sought to examine adaptation aftereffects along opposite and non-opposite morph trajectories. Here they implemented the paradigm of Rhodes and Jeffery (2006). In this paradigm, the opposite adapters corresponded to the anti-faces created from the morph trajectories of corresponding target faces when passed through an average face. While the non-opposite adapters corresponded to faces that did not follow along the same morph trajectory but laid an equal distance away from the target face as a matching anti-face would, while not passing through the average face. In this paradigm, if an exemplar-based model were to hold true, then there should be no difference between induced aftereffects of opposite and non-opposite adapters because this model substantiates that adaptation is a result of general perceptual biasing. Contrastingly, in a norm-based coding model, adapting to an opposite adapter will bias adaptation aftereffects, while non-opposite adapters will not because this model proposes that adaptation aftereffects are a result of opposite biasing. Results of Rhodes and Jeffery (2006) were supportive of a norm-based coding model, suggesting that adaptation likely biases in an opposite perceptual pattern. Within the simulation of Ross et al. (2014), only the exemplar-based and two-pool opponent models yielded similar results to Rhodes and Jeffery (2006). That is, when adapting to an opposite face, there was an increase in correct target identification, and when adapting to a non-opposite face there was less of an increase in correct target identification. The traditional norm-based coding model yielded inconsistent results, where adapting to both an opposite and non-opposite face increased correct target identification. Additionally, when examining different combinations of parameter values, exemplar-based and two-opponent pool coding models remained consistent regarding predictions, but no combinations of parameter values were able to produce accurate predictions for the traditional norm-based coding model. Meaning that in the simulations, the traditional norm-based coding model predicted there would be as much, if not more, adaptation for non-opposite conditions. These results for the traditional norm-based coding model run in contrast to the findings of Rhodes and Jeffery (2006). This would indicate that exemplar-based and a two-opponent pool coding models of face perception can predict an opposite bias, while a traditional norm-based coding model cannot. Together, the results of all these simulation paradigms suggest that facial adaptation aftereffects across multiple different paradigms and parameters can best be explained by both an exemplar-based and two-opponent pool coding. These findings are contrary to claims in face perception research that faces are likely represented in a norm-based coding manner.

Although it has been suggested that facial identity may possibly be coded in an exemplar-based manner, it has yet to be fully examined if facial expressions may also be coded in an exemplar-based fashion. However, support may come from evidence of behavioral, functional imaging, and computational studies showing that recognition of different basic expressions is based on signals from separate neural pathways. If facial expressions themselves are thought of as exemplars, then the absolute values for differing facial expressions may lead to the activation of different neuronal pools in different brain areas which aid in the facilitation of facial expression recognition. For example, regarding behavioral work it has been demonstrated that lesions to the amygdala can lead to significant impairments of fearful face recognition but not for other basic facial expressions (e.g., Adolphs et al., 1994). Therefore, it may be that the absolute values of key facial dimensions of fear lead to greater activation of neuronal pools within the amygdala than for other expressions. This may be due to other expressions not sharing homogenous facial features/dimensions with fear (e.g., eyes being more narrowed in a happy expression than in a fearful expression, where the eyes would be more opened (Ekman, 1971); thus, they do not activate the same or neighboring neuronal pools to as great of an extent.

Although most of the functional imaging studies examine differences of basic expressions compared to neutral, this does not support the same arguments of norm-based coding from the EEG results of Gwinn et al. (2018). This is because Gwinn et al. (2018) used neutral expressions as a baseline comparison to basic expression categories within a particular region of interest, while these later mentioned fMRI studies use neutral as a baseline across numerous brain areas. Thus, in these functional imaging studies, one can compare how basic expressions produce varying responses across different brain areas. One such fMRI study by Gorno-Tempini et al. (2001) examined differences in facial expression processing for disgust, happy, and neutral expressions. Results of their study demonstrated that when tasked with making explicit task judgments of emotion, disgust led to the activation of the right neostriatum and left amygdala, and happy judgments caused activation of the bilateral orbitofrontal cortex. Another fMRI study conducted by Sprengelmeyer et al. (1998) tasked participants with responding to the gender (i.e., either male or female) of a target face that displayed an expression from the emotional categories of either angry, disgust, fear, or neutral. This was done to examine the neural structures related to the expressions of anger, disgust, and fear compared to neutral. Results of the experiment showed that when processing angry faces, there was an increase in activity from the posterior regions of the right cingulate gyrus and the medial temporal gyrus of the left hemisphere. Additionally, when processing disgusted faces, there was greater activation in the right putamen as well as the left insula. Finally, for fearful faces, results showed increased activation of the right fusiform gyrus and the left dorsolateral frontal cortex.

Aside from functional imaging work, a recent computational study has also shown that expression recognition may occur in an exemplar-based manner. However, it should be noted that this study did not compare across other face coding models and face processing paradigms, such as the work done by Ross et al. (2014). Therefore, it is not yet known if an exemplar-based coding model may fit best when trying to explain how facial expressions are coded when compared to other models, such as a traditional norm-based or two-pool opponent coding model. Work by Farajzadeh and Hashemzadeh (2018) examined if a pool of simple and linear models of decoding were able to accurately predict what expression was being presented based upon prior knowledge of facial expressions (i.e., making predictions about the target facial expression based on already learned facial expressions within the classifier). This model is in line with an exemplar-based coding model because exemplar-based coding substantiates that facial recognition occurs based upon exemplars from other faces that were processed and perceived in the past and not through comparisons of a norm face held within the face-space. Their model was evaluated on five separate facial expression image databases, and the results showed that the model was able to accurately predict which emotional category the target expression belonged to when based on already learned facial expressions. It should also be noted that these high recognition rates of expression were found for all six basic emotional expressions of faces as well as neutral facial expressions.

CHAPTER 4. CLASSIFICATION OF RESPONSES TO FACES

Machine learning techniques have been used in conjunction with neuroimaging and other cognitive neuroscience methods to investigate numerous aspects of face perception. For example, multivariate pattern analysis (MVPA) of EEG signals can be utilized to examine if there is a potential nature of facial expressions within the brain, or eye-tracking decoding can be implemented to examine the potential differences in facial processing for special populations compared to neurotypical populations. The current chapter aims to provide an overview of research that has been conducted to investigate face processing while utilizing machine-learning methods. Specifically, this chapter will cover topics such as EEG decoding and eye-tracking decoding.

4.1. EEG Decoding

A technique known as EEG decoding may allow us to predict what facial expressions an individual is looking at by analyzing neural responses patterns obtained using EEG (e.g., as seen in Mares et al., 2021). Pattern classification of EEG signals has been used to examine how EEG signals are related to a variety of cognitive visual tasks. For example, neural responses to the orientation and location of a visual stimulus can be successfully decoded by EEG signals obtained while the information is stored in visual working memory (Bae & Luck, 2018). The decoding itself is based upon the patterns of observed activations of scalp activity recorded in EEG (Garcia et al., 2013). Thus, decoding may provide us with a glimpse into how the brain encodes complex visual stimuli or abstract semantic information (Tong & Pratte, 2012). Not only could it be postulated that one would be able to use EEG decoding to potentially predict what facial expression stimuli a participant was viewing, but it could also be used to investigate the influence of emotional valence on decoding performance surrounding classification accuracy and its temporal dynamics.

Previous research suggests that classifiers can predict what facial expression of the six basic expressions and neutral face stimuli participants are viewing (Greening et al., 2018; Liang et al., 2017; Nemrodov et al., 2019; Smith & Smith, 2019). Functional magnetic resonance imaging (fMRI) studies such as the ones conducted by Liang et al. (2017) and Greening et al. (2018) showed that fMRI alongside MVPA supports the above claim. To examine facial expression decoding in both motion- and face-selective brain regions. Participants were asked to view expressions of anger, disgust, fear, joy, sadness, and surprise in either a static image condition, video condition, or eye-obscured condition. The results of the study suggested that expressions are represented in both face- and motion-selective brain regions. (Liang et al., 2018). In the study conducted by Greening et al. (2018), they demonstrated that face-selective brain regions can be activated by partial face stimuli (e.g., displays of the eyes or mouth) regarding emotional expressions. Their results argued that the neural responses caused by the viewing of faces that displayed only the eyes or had the eyes removed allowed for reliable generalization of face-selective areas regarding these responses. Additionally, it was shown that the performance of the classifier was correlated to behavioral performance in both the Dorsal Prefrontal Cortex (dPFC) and Superior Temporal Sulcus (STS).

Not only has MVPA shown to work well alongside fMRI when it comes to the encoding/decoding of face stimuli, but research has demonstrated that EEG is also a reliable technique for examinations of face-related neurological responses. As stated earlier, Smith and Smith (2019) used MVPA and EEG to investigate how faces are processed across time regarding both facial identity and facial expression during explicit and incidental task conditions. Results showed that across the posterior electrodes, peak decoding occurred in the timeframe of 91 – 170ms during both the explicit and incidental tasks and for both facial identity and expression. Results also demonstrated that although peak decoding of facial identity was hindered during the incidental task, peak decoding of facial expression was not impacted by either task context. Taken together, these results suggest that the processing of facial expressions takes place to a larger degree than facial identity, especially when it occurs in incidental task conditions.

4.2. Eye-Tracking Decoding

During an experimental task, eye-tracking methods can be utilized to record saccades, fixations, and temporal information of such eye movements. This information can be useful when aiming to investigate allocations of visual attention and any potential differences between diagnostic and control groups (Carter & Luke, 2020). These eye-related signals can then be combined with machine learning approaches to investigate any potential differences in various aspects of face processing (e.g., time spent fixating on the face vs. fixating on the background of an image).

A recent study by Kang et al. (2020) sought to examine if they could potentially identify children with ASD by combining the machine learning approach of using a support vector machine (SVM) with EEG and eye-tracking-related signals. In this study, 97 children ranging from the ages of three to six were first hooked up to an EEG system as resting-state data was collected. During resting-state EEG data collection, participants were tasked with sitting in a chair in a shielded room and asked to keep quiet while minimizing blinks and bodily movements for six minutes. Following this, the children performed an eye-tracking task while being exposed to neutral faces that either belonged to a member of their own race or a member of another race. In this eye-tracking task, the areas of interest (AOI) selected for analysis include the background, body, face, eyes, right eye, left eye, mouth, and nose of the stranger's face. The data from the EEG and eye-tracking methods were then combined and plugged into an SVM to classify for autistic vs. typically developing children.

Results showed that the classifier was able to decode significantly above chance level (i.e., 50% for the two categories of autistic or typically developing) with a maximum classification accuracy reached at 85.44%. Additionally, the classification accuracy for combining the EEG data and eye-tracking data was higher than when decoding for both sets of data separately. For example, the researchers found that when only using EEG resting-state activity for the data plugged into the SVM, they obtained a classification accuracy of 68%, and when only using the eye-tracking data, they reached a classification accuracy of 72.33%. These results would suggest that combining these two methods may potentially lead to higher decoding accuracy when classifying for autistic and typically developing individuals, and that even separately, the data from these two methods can lead to above-chance decoding for these groups. However, some limitations of this study should be addressed. For instance, this study only used neutral expressions for their stimuli presentations. It may be possible that there could be potential differences found between autistic and NT individuals when concerned with classification scores for multiple expression categories (e.g., angry, happy, and neutral). Additionally, this study only examined data from children, so it is not yet clear if there are any potential differences in classification performance for autistic and typically developing adults. Finally, the EEG resting-state data was not representative of any potential differences in face processing between autistic and typically developing individuals.

A study by Jiang et al. (2019) did not have the limitations listed for the prior mentioned study and used a different machine-learning approach for analyzing eye-tracking data. This study examined the eye-tracking data of 23 subjects with ASD and 35 control individuals using a Dynamic Affect Recognition Evaluation (DARE) task to investigate atypical gaze patterns of individuals with ASD. This task required participants to view dynamic videos of angry, disgusted, fearful, happy, sad, and surprised faces while having them categorize which emotional category the face belonged to. Behavioral and eye-tracking data were classified using random-forest (RF), a machine-learning method that creates a forest of decision trees for classification. The RF was trained to classify the fixations of the ASD and NT groups, then trial-level and subject-level classification scores were obtained by averaging all fixations of the same trial or subject. Results showed that a combination of task, gaze, and face features allowed for classification accuracy as high as 72.5%, and 75.6% at the trial level and 86.2% at the subject level. These results further highlight the feasibility of using eye-tracking data to accurately decode whether an individual potentially has autism or not. However, one limitation of this study that should be noted is the use of lexical prompts to categorize which expression best described the face while not also controlling for levels of alexithymia in these individuals (Refer to **Chapter 5** for more information on alexithymia).

`CHAPTER 5. DISRUPTIONS/DEFICTIS OF FACE PROCESSING AND PERCEPTION

Face processing and perception within typically developing individuals can be disrupted in a myriad of manners, leading to deficits in the recognition of facial expressions and identities. An example of this can be found in the face-inversion effect (Yin, 1969), where individuals study the identities of upright faces and upright complex inanimate objects (e.g., houses) and are then tasked with identifying a series of face and object stimuli that are instead presented in an inverted manner (upside down). Results consistently show that individuals perform significantly worse at recognizing inverted faces compared to inverted inanimate objects, suggesting that inversion likely disrupts the holistic processing of faces to a larger degree than the holistic processing of objects. Additionally, outside of facial identity, it has also been shown that when typically developing individuals are unable to holistically process facial expressions, they also express decrements in behavioral responses, such as response times.

A study conducted by Wang et al. (2014) examined the responses of amygdala neuronal units regarding image properties of emotional face stimuli, perceptual judgments of emotion, and motor outputs surrounding behavioral categorizations of emotions. Outside of these focal areas of examination, an important finding of their study was that individuals, whether they be NT controls or epileptic patients, did not exhibit a PCA while performing a "bubbles" task. During this task, epileptic and control individuals were asked to categorize presented facial expressions of happy and fearful stimuli that had randomly selected parts of the face to be visible while the rest of the face was occluded. The behavioral results indicated that there was no significant difference between the response times for the epileptic and control individuals, nor either emotion category. These results demonstrated that the PCA could be diminished within typically developing and patient population individuals.

The above results indicate that disruptions of face processing in typically developing individuals occurs in viewing conditions such as face inversion and occlusion. However, there are consistent reports suggesting that face processing, perception, and judgments can be disrupted by developmental disorders irrespective of viewing condition. These studies suggest that certain atypically developing individuals may process and perceive faces differently from typically developing individuals. The current chapter aims to cover topics such as disruptions and deficits of face processing and perception amongst atypically developing individuals belonging to ASD and alexithymia groups. These will be key areas of investigation for the presently proposed study.

5.1. Face Processing and Perception in Individuals with Autism Spectrum Disorder

The diagnostic criteria of the DSM-5 state the following must be met for a diagnosis of ASD to be made. Individuals must exhibit “persistent deficits in social communication and social interaction across multiple contexts,” display “restricted, repetitive patterns of behavior, interests, or activities,” that “symptoms must be present in early childhood (but may not fully manifest until social demands exceed limited capacities),” and that “symptoms together limit and impair everyday functioning”

(American Psychiatric Association, 2013). Additionally, research has indicated that regarding the processing and perception of faces, individuals with ASD are likely to exhibit differences in eye-gaze patterns (e.g., Kang et al., 2020), a restriction of their ratings on the perceived valence of facial expressions (e.g., Tseng et al., 2014), potential difficulties in expression labeling (e.g., Griffiths et al., 2019), and neurological differences of brain regions and responses implicated in face processing and perception (e.g., Van der Donck et al., 2019).

5.1.1. Eye Gaze

As stated in the previous chapter, the study conducted by Kang et al. (2020) sought to combine EEG and eye-tracking methodologies and use measurements from both modalities to decode for which children may or may not be autistic. This study was comprised of an ASD group and a control group, both of which had ages that ranged from three to six years old. The researchers first grabbed resting-state EEG activity from each participant for six minutes. Following this recording, participants were asked to partake in an eye-tracking task, where participants were shown a series of neutral female face stimuli that either belonged to their own race or another race. Areas of interest (AOI) from the face images used within this study were the background of the photo, and the body, face, eyes, right eye, left eye, mouth, and nose of the actor in the picture. Comparisons of eye-tracking data from ASD and NT control children revealed significant differences amongst the groups regarding fixation duration percentages for background, face, mouth, and nose. Results showed that compared to NT controls, the ASD children gazed more at the background and less at the face, nose, and mouth of an own-race face. Additionally, it was shown that when viewing another-race face, ASD children gazed more at the background and less at the nose and mouth when compared to the NT control children. These results suggest that concerning face processing, there are likely differences in eye-gaze patterns amongst ASD and NT individuals and that these differences can be detected as early as three to six years of age. However, it should be noted that this study has its limitations. For instance, the researchers did not have face stimuli generated from male actors or stimuli from multiple expression categories.

One study which did examine differences in eye gaze patterns surrounding emotional facial expressions was conducted by Kliemann et al. (2012). Eye-tracking and fMRI methodologies were combined to investigate if dysfunctional amygdala response patterns in ASD support an active avoidance of direct eye contact or a lack of social attention. There were 17 NT adult controls and 16 ASD adults who partook in an emotion classification task while seated in a scanner with an embedded eye tracker. In the emotion classification task, participants first had to fixate on a fixation cross located centrally on the presentation screen for 2 – 5.15 s, then were shown a facial expression for 150 ms that belonged to an emotional category of either happy, fearful, or neutral, followed by a blank screen for 2 – 14 s, and finally a response screen for 2 s in which they would indicate which expression was shown via button press (i.e., 1 for happy, 2 for fearful, 3 for neutral). To examine the influence of initial fixation, faces were shifted vertically, either upwards or downwards, so the mouth or eyes would be where the previously presented fixation cross was located. Eye-tracking results revealed that there was a marginally significant difference between ASD and NT adult groups for neutral faces when initially fixating on the mouth region. Additionally, it was shown that when initial fixation occurred at the mouth region, the NT adults showed a greater amount of fixation changes from the mouth to the eyes than the ASD adult group and that the ASD group was more likely to move their eyes away from the eyes of the presented face stimuli. These results were also found across all expressions and initial fixation locations, however, there was only a significant difference amongst the neutral faces. Finally, these eye movement patterns were also correlated with the number of correct responses in the emotion recognition task in the ASD but not the NT group. This group difference was significant for trials in which the initial fixation occurred at the mouth region, and marginally significant for all trials. These results would indicate that more participants gazed away from the eyes when a lesser number of correct responses were made, especially when fixations initially occurred at the mouth region.

Although there have been conflicting results as to whether ASD children exhibit facial expression recognition difficulties, work from Fridenson-Hayo et al. (2016) would suggest that these difficulties occur cross-culturally, suggesting that there is a universal nature of facial expression recognition difficulties in ASD. To better understand visual processing strategies that may relate to facial expression recognition difficulties observed in ASD and the developmental trajectory of these difficulties, Black et al. (2017) systematically reviewed studies that exposed ASD individuals to eye-tracking methods. In total 33 eye-tracking studies and 1 eye-tracking and EEG study were included in their review. Results were then analyzed according to the age group of the participants (i.e., children, adolescents, and adults) and the stimuli used (i.e., static basic emotions, static complex emotions, dynamic basic emotions, and dynamic complex emotions).

A reduction of eye gaze is commonly considered a key characteristic of ASD according to the American Psychiatric Association (2013), however, results of the eye-tracking studies included in Black et al. (2017) were inconsistent. Of the nine child studies included in their review, only two found a significant difference in gaze patterns of ASD individuals and NT individuals. Nuske et al. (2014a) and Nuske et al. (2014b) had found that ASD children, compared to NT children, displayed significantly fewer fixations and dwell times for the eyes when exposed to either static or dynamic presentations of fearful or neutral faces. Nuske et al. (2014b) found that children exhibited significantly shorter fixations to the eyes of static fearful and neutral faces across all stimulus presentation times (i.e., 30 ms, 300 ms, 2000 ms). Additionally, they observed that ASD children made shorter fixations to the mouths of static neutral faces but not to static fearful faces. For dynamic presentations of fearful and neutral faces, Nuske et al. (2014a) found that ASD children only displayed eye aversion behaviors for neutral faces and not fearful faces. Regarding adolescent results, only four of the eight included studies suggested there may be an increase in eye aversion behaviors in ASD adolescents compared to NT adolescents. Dalton et al. (2005) showed that adolescents with ASD exhibited significantly fewer fixations to the eyes of static presentations of angry, fearful, happy, and neutral faces. Pertaining to results of static complex emotions, Hanley et al. (2013) showed that adolescents with ASD looked at the AOI of hair for naturalistic isolated faces (i.e., faces taken from a clip of actors involved in a non-scripted social interaction) compared to NT adolescents, but no significant differences for the AOIs of eyes, nose, mouth, face, body, background and other for the naturalistic isolated faces. Tottenham et al. (2014) showed that adolescents with ASD, compared to NT adolescents, were significantly more likely to look away from the eyes of static neutral faces that they perceived as more threatening, and that this behavior was not observed for static angry expressions. White et al. (2015) found that there were no significant differences in fixations to the eyes of faces; however, it was found that adolescents with ASD exhibited a shorter fixation time to the eyes when controlling for self-reported measures for fear of being judged negatively. Finally, of the 16 adult ASD studies, 11 observed increased eye aversions in ASD adults compared to NT adults regardless of expression, if participants completed a free viewing or active recognition task, or if the expressions were presented statically or dynamically. Together, these results suggest that the differences in eye gaze patterns for ASD and NT individuals observed during facial expression processing may be more reliably found as ASD individuals get older. Black et al. (2017) suggest that this may be due to ASD individuals experiencing differential development of basic emotional face processing in late childhood or adolescence, whereas basic emotional face processing for NT individuals reaches maturity in late childhood.

5.1.2. Restricted Valence and Arousal Ratings

A study conducted by Tseng et al. (2014) examined potential differences in valence and arousal ratings for individuals with autism and typically developing controls. Previous work has been conflicting as to whether individuals with ASD exhibit disrupted or normal autonomic responses to emotional content. For example, results of Ben-Shalom et al. (2006) showed that skin conductance responses to pleasant, unpleasant, and neutral images did not significantly differ between high-functioning ASD children and control children. Blair (1999) showed similar results for low-functioning ASD children except for threatening stimuli, where it was demonstrated that they were hypersensitive to threatening images compared to the control children. Tseng et al. (2014) suggest that another potential reason for inconsistencies in the emotion recognition literature of autism may be due to the model of emotion used in many studies. Other studies have adopted a traditional Theory of Basic Emotions or discrete emotion theory that suggests that basic emotions are the product of distinct neural systems that display discrete patterns of motor behaviors, autonomic responses, and facial expressions. They argue that this model can introduce complications, such as difficulty in disentangling the influence of arousal and/or valence on expression recognition abilities. Instead, the authors propose that it would be best to adopt the Circumplex Model of Affect. This model proposes that emotions are extracted from two orthogonal dimensions of emotional experience, valence, and arousal. In this model, the dimension of valence is the degree of pleasantness that describes the emotion (pleasant or unpleasant), and arousal describes the energy that describes the emotion (high or low energy). Additionally, this model of affect posits that a linear combination of arousal and valence can be used to represent all emotions.

To investigate potential differences in ratings of arousal and valence, Tseng et al. (2014) subjected participants to an affective circumplex task, wherein, participants were shown emotional face images while placed in an fMRI scanner and asked to simultaneously rate the valence and arousal of presented face images via mouse click on a 9 x 92-dimensional grid. The right side of the 9 x 92 grid held the dimension of “High Pleasure” (i.e., valence) and the top of the grid held the dimension of “High Energy” (i.e., arousal). For example, if a participant made a response on the far right and towards the top of the grid, they would be responding in a way where they would have rated the face as having high levels of pleasantness (valence) and energy (arousal). Conversely, if they made a response towards the far left and lower half of the grid, they would be rating the face as having low levels of pleasantness (valence) and energy (arousal). The expressions to which the participants were exposed to consisted of angry, bored, contented, disgusted, fearful, happy, neutral sad, scared, sleepy, and surprised faces. For their analyses, participants were divided into four distinct groups, adults with autism, typically developing adults, children with autism, and typically developing children.

Their results found a significant main effect of diagnosis, wherein individuals with autism were constricted for the entire circumplex of their arousal and valence ratings. This suggests that individuals with autism likely do not rate valence and arousal scores as high as their typically developing counterparts. Regarding results of exploratory analyses of emotion-specific effects, the authors found that ASD children rated high arousal emotions, such as excitement and surprise, significantly lower than NT adults, and they rated low arousal emotions, such as bored and sleepy, as significantly more arousing compared to NT adults. ASD children also rated negatively valenced emotions, such as anger and sadness, as significantly less positive, and the positively valenced emotions of excitement and happiness were rated significantly less positive compared to ratings from the NT adults. Regarding emotion-specific comparisons between ASD adults and NT adults, it was found that negatively valenced emotions of anger, fear and surprise were rated significantly less arousing in the ASD adults compared to the NT adults. Additionally, the negatively valenced emotion of sadness was rated as more positively valenced in the ASD adults compared to the NT adults. Results also showed a significant constriction in arousal ratings for the positively valenced emotions of excitement and happiness for ASD children compared to NT children. Finally, there were no significant changes in rating over development in the ASD group, but NT adults displayed significantly higher arousal ratings than NT children for the negatively valenced emotion of anger and a significantly less positive valence rating for the high arousal emotion of excitement.

These findings are consistent with previous results that have shown restricted valence and arousal ratings when tasking ASD and NT participants to view images from the International Affective Picture System (IAPS) (Ben-Shalom et al., 2006; Lang et al. 1999). The authors also suggest that these results may provide an argument for how in prior studies, individuals with autism are more likely to rate “exaggerated” expressions (i.e., an emotional image created by increasing its intensity relative to the average emotional expression category) as more realistic. It should also be noted that when examining task performance, ASD and NT groups used the full range of scores. This suggests that results were not due to only the NT groups using the full range of scores and that all groups were performing the task correctly. Finally, the authors found no significant main effect for age or sex. Together, these results would suggest that individuals with autism are restricted in their emotional ratings of facial expressions and that these differences can be found in multiple demographic categories of those with autism, such as in children vs. adults and males vs. females.

5.1.3. Potential Deficits in Expression Recognition

As previously mentioned, individuals with autism show impairments in social communication and interaction and display difficulties in identifying the emotional expressions portrayed by others. However, results remain inconsistent when investigating potential deficits in identifying emotional facial expressions for individuals with autism.

Griffiths et al. (2019) suggest that these differences may stem from some studies only using high-intensity expression facial stimuli during an expression recognition task. They suggest that individuals with ASD may use compensatory strategies rather than automatic affective processing to recognize expressions and that these compensatory strategies may fail when task demands are too great. For instance, the authors discuss how impairments in recognition have been shown in instances with lower intensity, which is in contrast with results displaying no deficits in recognition for high-intensity expressions.

To examine potential differences among ASD and NT individuals at multiple expression intensities, Griffiths et al. (2019) subjected ASD and NT children to an Emotion Recognition Task (ERT). They hypothesized that ASD individuals would demonstrate impairments in recognition for the six basic facial expressions compared to NT individuals, and that emotion recognition ability would be correlated with parent-reported social skills in the ASD group. The emotion recognition task was comprised of the six basic emotional expressions at eight varying intensity levels. Stimuli were generated by creating four different prototypes (i.e., male adult, female adult, male child, female child) for each emotional category (i.e., creating an average amongst 12 – 15 individuals of the same age, gender, and expression), and by generating an emotionally ambiguous expression for each prototype by averaging all expression images. Participants were then exposed to an 8-step morph sequence of expressions that began from the ambiguous expression up to a full-intensity emotional expression for each prototype. During the experiment, participants were only exposed to the stimulus for 300 ms, as to minimize the use of any compensatory strategies. Following stimulus exposure, participants were presented with a visual mask for 150 ms, followed by a forced-choice response screen in which they would select which of the six basic emotional expressions the presented stimulus belonged to.

Results showed that the ASD participants were significantly less accurate than controls at recognizing expressions across both high and subtle expression intensities. However, as intensities reached a very low level, there were no significant differences between the ASD and NT groups, which may have been due to floor effects. When collapsing across the six-highest intensity levels and comparing group performance for each emotional expression, results showed significant group differences for the expressions of anger, disgust, happiness, and sadness, and that surprise and fear emotional expressions were approaching a significant difference between the two groups. Regarding confusions or errors made, results showed that there was no significant difference between the groups. These results run contrary to arguments that there may be a negative-response bias in ASD individuals and may instead suggest that ASD individuals likely use the same emotional cues to categorize expressions but to a less accurate degree than their NT counterparts. Altogether, these results suggest that ASD individuals are more likely to express a deficit in expression labeling compared to NT individuals and that these deficits can be seen across various intensity levels. However, it should be noted that this study, like many other ASD studies, used emotional labels for expression categorization. Due to the high rate of co-occurring alexithymia in ASD individuals, it may be best to move away from lexical labels of expression categorization (Refer to **Chapter 5.2** for more information on alexithymia).

Leung et al. (2018) argue that inconsistencies in face processing ASD research may be due to numerous differences found between studies surrounding their inclusion and exclusion criteria or the paradigms used during experiments. For example, differences in the samples used as seen in the ages included (e.g., study A only examines children, while study B examines children and adults), differences in their dependent variables (e.g., study A examines 3 expressions, while study B examines 7), and/or incongruencies of experimental procedures (e.g., duration of exposure to stimuli). Additionally, regarding the duration of stimulus exposure, it is suggested that ASD individuals may use compensatory strategies during facial expression recognition, then it could be argued that emotion recognition in autism may not be automatic. Thus, if a study exposes ASD participants to a facial expression stimulus for too long, then there is a chance that compensatory mechanisms are implemented, which may lead to inconsistent findings of potential differences in facial expression processing. Regarding the limitation of expressions used within ASD research, many have chosen to use the expression of fear to examine potential differences. However, an argument can be made that the negatively valenced expression of anger may be better to use when investigating potential differences in facial expression processing. This is because anger elicits a stronger avoidance response, is more likely to be encountered in daily life, and is expressed when a violation of social norms occurs, thus making it potentially more ecologically valid. Fear on the other hand elicits a stronger approach response and when expressed may play a role in appeasement or submission (Marsh et al., 2005).

Prior to the work of Leung et al. (2018), no study had yet examined the temporal neural correlates of expression processing related to anger. To minimize this limitation, they investigated the neural mechanisms of implicit expression processing related to the expressions of angry, happy, and neutral as MEG and MRI data were acquired.

Participants were comprised of adults that were equally split between those with ASD and controls. Those belonging to the ASD group completed the Autism Diagnostic Observation Schedule (ADOS-G, ADOS-2) and had their diagnosis confirmed by an expert clinician. Task procedures required participants to view an emotional facial expression that was presented simultaneously alongside a scrambled pattern (the scrambled image of the face). Face stimuli consisted of seventy-five faces (25 for each emotional expression portrayed by 13 male and 12 female actors). Participants completed 300 trials (50 trials of each expression in each hemifield), with each face being presented for 80 ms twice in each hemifield and were asked to indicate the location of the scrambled image.

Behavioral results indicated no significant between-group effects on accuracy or response latency for either the ASD or NT groups. Between-group neuroimaging results concerning the implicit processing of angry faces indicated greater activity in the ASD group than controls in the right inferior temporal, left fusiform, right fusiform 425 ms, and the right amygdala/fusiform, and greater prolonged activation in the right anterior insula. Additionally, it was shown that there was decreased activity in the ASD group compared to the control group in the occipital areas of the left cuneus, left middle occipital gyrus, left precuneus, and right postcentral/supramarginal gyrus, and less activity was also observed in the left caudate and left anterior cingulate cortex (ACC). Between-group neuroimaging results concerning the implicit processing of angry faces indicated greater activity in the ASD group than controls in the left inferior temporal gyrus, left fusiform gyrus, right inferior temporal gyrus, right fusiform gyrus, right anterior cingulate, and anterior insula. It was also shown that there was decreased activity in the ASD group compared to the control group in the left middle occipital gyrus, left precuneus, left cuneus, right postcentral/supramarginal, left caudate, and left ACC. Within-group comparisons revealed no significant differences in activation for happy and angry faces in the ASD group. However, within-group comparisons for controls revealed that activation for angry faces was significantly greater than for happy faces in the right inferior occipital, the right calcarine, and right postcentral gyri.

Together, these results suggest there are likely atypical neural activations for angry and happy faces in adults with ASD despite comparable behavioral performance. The temporal results of this study would suggest that limbic and occipital regions in ASD individuals are differentially recruited as early as 100 ms post-stimulus presentation for both angry and happy faces. Results also suggest that individuals with ASD may exhibit increased activation of the anterior insula, fusiform gyrus, inferior temporal gyrus, ACC, and amygdala, which are thought to be key areas for emotional processing. Additionally, ASD individuals not exhibiting differential activation for happy and angry faces as compared to controls that did exhibit differential neural activity further suggests that ASD individuals likely process emotional facial expressions differently than NT individuals.

5.2. Face Processing and Perception in Individuals with Alexithymia

Alexithymia loosely translates from Greek to English as "no words for emotions" and is defined as a personality trait that is marked by a reduced ability to identify and describe internal emotional experiences and the emotional experiences of others (Lane et al., 1996; Ola & Gullon-Scott, 2020). Additionally, it has been suggested that inconsistencies in findings related to possible deficits of facial expression recognition abilities seen within autistic individuals may instead be better described by frequently co-occurring alexithymia. Meaning, that autistic individuals who do exhibit a deficit in facial expression recognition ability may be experiencing these difficulties in relation to alexithymia itself and not autism (Bird & Cook, 2013). Currently, it is estimated that roughly half of the autistic population exhibit levels of alexithymia, while 10% of the typically developing population display levels of alexithymia. Additionally, it has been shown that individuals with higher levels of alexithymia, whether autistic or typically developing, exhibit a difference in eye gaze patterns and a reduction in recognition abilities surrounding facial expression processing (Bothe et al., 2019).

5.2.1. Eye Gaze

As stated earlier, autistic individuals are less likely to attend to the face region during face-related tasks, particularly the eye region during a facial expression recognition task (Kang et al., 2020; Kliemann et al., 2012). One study that examined the influence of levels of alexithymia on visual attention while identifying facial emotions was conducted by Fujiwara (2018). In this study, selection criteria were if individuals exhibited high or low levels of alexithymia based on Toronto Alexithymia Scales (TAS-20) scores, and visual attention was measured by eye movement patterns as participants were exposed to either relatively clear or ambiguous blends of facial expressions. These blends were used because it is believed that ambiguous facial expressions may be more difficult for alexithymic individuals to recognize. The authors predicted that individuals with high levels of alexithymia would perform worse at facial expression recognition than individuals with low levels of alexithymia and that the high alexithymic individuals would perform particularly worse when exposed to the ambiguous blends. Additionally, they also predicted that there would be a reduced eye preference for the high alexithymic individuals as they processed the facial expressions.

In this study, participants completed the emotion hexagon – identification of mix-ratios in expression continua task. Participants were asked to judge the mixture ratios of two emotional expressions blended into one facial image. The blended faces had a dominant emotion of at least 55% and at most 95%, and blends increased in 10% increments. Clear blends were defined as having a mixture ratio of 85:15 or 95:5. Ambiguous blends were defined as having a mixture ratio of 55:45, 65:35, or 75:25. On-screen, above the presented blended image were two reference faces that belonged to the same identity and expressions used in the blend. To judge the mixture ratios of the blends, participants clicked via a mouse on a visual analog scale (e.g., selecting directly in the middle of the scale would indicate a response of a 50-50 mixture ratio). As participants completed the task, eye movements were recorded via an eye tracker. AOI were defined as the three faces and the eye and mouth regions of each of the three faces. Dwell times within each AOI were required to have a minimum duration of 60 ms.

Concerning accuracy, results showed a main effect of ambiguity and an interaction between ambiguity and emotion. However, contrary to the hypothesis, there was no main effect or interaction observed for alexithymia scores. Meaning, that regardless of ambiguity, individuals with high levels of alexithymia did not perform significantly worse than individuals with low levels of alexithymia. Results also showed a significant main effect of group for eye preference, where low alexithymic participants exhibited a stronger eye preference. Indicating, that low alexithymic individuals spent more time attending to the eye region than high alexithymic individuals. Additionally, eye movement results indicated an interaction for ambiguity and group. It was shown that the high alexithymic individuals had a smaller eye preference in ambiguous trials compared to clear trials, and that eye preference for low alexithymic individuals did not vary across ambiguity levels. Finally, hierarchical regression analysis was used to look at the relationship between deviance scores (inverse accuracy), eye movements, and alexithymia. A significant model was only found for the Alexithymia x Eye Movement interactions. In the high alexithymic group, higher deviance scores were observed when there was an increase in eye preference of the target face, and the opposite relationship was seen in the low alexithymic group. Additionally, in the high alexithymic group, there were higher deviance scores when these individuals exhibited smaller dwell times on the reference faces, and the opposite was found in the low alexithymic group, where they performed worse when dwell times were longer for the reference faces. Together these results may suggest a compensatory mechanism in eye gaze patterns in the high alexithymia group, enabling them to have accurate judgment scores compared to the low alexithymia group, and that there may be reduced attention to eye regions of the face in relation to having alexithymia. However, some limitations of this study should be addressed. Even though a reduced preference for eye regions has been demonstrated for alexithymic individuals in this study and autistic individuals, as seen in other studies, Fujiwara (2018) did not control for if participants had autism or not. Therefore, the degree of similarity between these two groups regarding eye movement patterns during facial expression recognition could not be observed.

5.2.2. Deficits in Expression Recognition

Lane et al. (1996) examined if the observed emotional recognition deficits in individuals with alexithymia are limited to only verbal emotional responses. They argued that if alexithymia is a consequence of impaired symbolic representation of emotion and that if experimental conditions measure emotion recognition ability via the utilization of stimuli or responses which involve an emotional word, then poor performance may instead be linked to an inability to comprehend and use the emotional terms appropriately. In Lane et al. (1996), participants were asked to complete two independent measurements of alexithymia (the Levels of Emotional Awareness Scale (LEAS) and the TAS-20) and a Perception of Affect Task (PAT). Based on the measures for alexithymia, participants were divided into three distinct groups for later analyses. These groups were individuals with alexithymia, intermediates, and individuals without alexithymia.

The PAT was a 140-item task that asked participants to match a verbal or nonverbal response to verbal or nonverbal stimuli across four unique subtasks that were each comprised of 35 items. Across each subtask, five sets of stimuli of the six basic expressions and neutral were presented (5 sets of stimuli x 7 expressions = 35 items), and participants were tasked with choosing the correct response from a display of seven potential responses that corresponded to each of the seven emotions relating to the stimuli. Subtask 1 was a verbal-verbal response task in which participants were exposed to sentences that did not contain any emotional words yet depicted a specific emotion, and then had to select which emotional word best described the sentence. Subtask 2 was a nonverbal-verbal response task, where participants were shown emotional face images and had to select which emotion best described the presented facial expression. Subtask 3 was a verbal-nonverbal response task where participants again read sentences depicting emotion but now had to select which facial expression belonged to the sentence. Lastly, subtask 4 was a nonverbal-nonverbal response task, where participants were shown emotional facial expressions and had to respond by indicating what emotional scene (that did not contain any faces) may best relate to the facial expression.

Results showed significant correlations between higher TAS-20 scores and demographic variables, such as lower SES, male gender, and older age. Correlations were also found between lower LEAS scores and lower PAT scores. These results would suggest a need to control for demographic variables when investigating the potential influence of alexithymia on emotion recognition abilities. Additionally, performance on PAT between each level of alexithymia was significantly different. Post hoc analysis revealed that accuracy rates for alexithymic individuals were significantly lower than those of non-alexithymic individuals across all subtasks. Additionally, differences between alexithymic and intermediate individuals were significant across all subtasks except for subtask 2 (faces-words; nonverbal-verbal), which was approaching significance. Together, these results may suggest that individuals with alexithymia have a greater impairment for emotional information processing, not just for facial expressions, and this may be what leads to the characteristic trait found in alexithymia, which leads these individuals to have a greater difficulty putting emotions into words.

Another more recent study that showed the importance of controlling for demographic information when examining facial expression recognition abilities was conducted by Ola and Gullon-Scott (2020). The authors discuss how inconsistencies in prior literature on autistic individuals demonstrating a deficit in facial expression recognition abilities may be due to not controlling for frequently co-occurring alexithymia. And that prior to the time of writing, all but one study investigating the alexithymia hypothesis included only males in their study (e.g., Bird et al., 2013). Meaning, that any prior results found relating to the role of alexithymia in facial expression recognition abilities had been conducted on male-only samples. Thus, previous findings on alexithymia and ASD may not generalize well to females with ASD and/or alexithymia. To address this limitation, the researchers aimed to include an all-female ASD sample and examined the role of alexithymia as they completed a more ecologically valid emotional facial expression recognition task, the Geneva Emotion Recognition Test-Short (GERT-S). The authors hypothesized that there would be a significant negative correlation between alexithymia and facial expression recognition abilities and that there would be no association between autism severity and facial expression recognition abilities.

Participants in this study were comprised of 83 females who stated that they had a clinical diagnosis of autism, ranging in age from 19 to 65, with a mean age of 38.5. Participants were asked to complete the Autism Spectrum Quotient (AQ) and TAS-20 for measurements of autism severity and alexithymia levels, respectively, before completing the GERT-S. The GERT-S was comprised of 42 video clips that could last from 1 – 3 s and displayed one of 10 actors (five males and five females) expressing one of 14 unique emotions (anger, irritation, disgust, sadness, despair, fear, anxiety, surprise, interest, relief, pleasure, amusement, joy, and pride). The actors in the clips had their upper torso and face shown as they portrayed an emotional expression using facial expressions, gestures, and vocalizations (pronunciations of phrases with no semantic meaning). After the stimulus presentation, participants were shown an emotional wheel that arranged each of the 14 possible emotional responses in a circle, and participants were forced to make a judgment on which emotion best described the stimulus.

Results showed that over 80% of the sample scored above the clinical cutoff value on the AQ and that based on TAS-20 scores, 72.3% of the sample could be categorized as alexithymic, 15.7% as borderline, and 12% as not having alexithymia. All further analyses had TAS-20 and AQ scores on a continuous scale, with participants having higher or lower levels of alexithymia and autism severity. Regarding GERT-S performance, results showed that participants responded correctly on average 54.1% of the time and that the average response time for correct trials was 4400ms, which was significantly shorter than incorrect trials at 6482 ms until a response was made. When examining the alexithymia hypothesis, the authors found a significant positive relationship between TAS-20 and AQ scores, such that greater autism severity was related to higher levels of alexithymia. Additionally, consistent with the alexithymia hypothesis, the authors found no relationship between AQ Total scores and accuracy on the GERT-S and a significant negative correlation between TAS-20 and GERT-S accuracy. These results would suggest that higher levels of alexithymia were better associated with worse facial expression recognition abilities compared to autism severity scores. Regarding subscales within the TAS-20 and AQ, results showed the following, higher TAS-20 scores on Difficulty Identifying Feelings and Externally Orientated Thinking were significantly associated with poorer accuracy, and AQ communication was the only AQ subscale significantly associated with poorer facial expression recognition abilities.

To examine whether TAS-20 predicted facial expression recognition abilities above and beyond autism severity, hierarchical regression analyses were conducted. Results showed that AQ Total scores were not a significant predictor of facial expression recognition abilities, but TAS-20 scores were. Additionally, multiple regression analysis showed that the AQ Total and TAS-20 Total scores were not significant predictors of average response time. These results run contrary to the alexithymia hypothesis by suggesting that alexithymia did not significantly reduce emotional processing speed in comparison to autism severity. To further substantiate their results, the authors conducted the same analyses on a separate sample of females with significant clinical levels of autism and who self-identified as having autism. In this separate sample, the same pattern of results for alexithymia and facial expression recognition abilities, as well as response times, were found. However, now AQ Total, AQ Communication, and AQ Imagination were significantly correlated with facial expression recognition abilities, but after multiple regression analyses, only alexithymia remained as a significant predictor of facial expression recognition abilities. Together, these results would suggest that facial expression recognition abilities in autistic adult females correlates with alexithymia and not autism. These results also further highlight the importance of controlling for demographic information when investigating the alexithymia hypothesis.

To better understand the neural contributions of alexithymia vs. autism in facial expression processing, Desai et al. (2019) used EEG methods to investigate potential differences in the structural encoding of faces and recognition of emotions. The ERPs they investigated were the P100 because it is suggested to be reflective of basic visual percepts, the N170 due to being implicated in the structural encoding of faces and potentially being representative of emotional face information, and the N250 due to it being reflective of higher-order decoding of emotional face information. Previous studies that have investigated potential differences in P100 and N170 responses of ASD and NT individuals have shown that ASD individuals have longer N170 latencies for faces (e.g., Kang et al., 2018) and that ASD individuals do not show an increase in P100 amplitude for inverted faces compared to upright faces (e.g., Hileman et al., 2011). However, prior to the work of Desai et al. (2019), no study had yet investigated ERPs related to face processing in alexithymia. The authors hypothesized that the influence of the ASD on the N170 and P100 would be consistent with results from previous literature and that alexithymia would influence the N250 due to the N250 being associated with higher-order decoding of emotions.

Participants of this study included 27 NT adults between the ages of 19 and 28. Levels of autism and alexithymia were measured via the AQ and the Bermond Vorst Alexithymia Questionnaire (BVAQ), respectively. Participants were then asked to complete an emotional expression passive viewing task that only required them to make a response via button press each time they saw a grey ball on the screen. Face stimuli consisted of 240 dynamic, grey-scale, computer-generated faces comprised of 210 unique faces. Of the unique faces, 70 were fearful, 70 were neutral (with a subset having puffed cheeks), and 70 were biologically impossible faces that displayed either the eyes or the mouth towards the scalp or chin. Participants completed five blocks of the experiment, and each block was comprised of 14 fearful faces, 14 neutral faces, and 56 faces that were to be excluded from future analyses (i.e., the biologically impossible or puffed-cheek neutral faces).

Behavioral results showed that the AQ and BVAQ scores were not significantly correlated, suggesting that they measured different attributes. ERP results of P100 amplitude and latency in response to emotion found no significant effects or interactions for either AQ or BVAQ scores. Suggesting that neither alexithymic nor autistic traits significantly influenced early perceptual processing of faces. Regarding N170 amplitude, results showed no significant main effect of either emotion or hemisphere on N170 for either AQ or BVAQ scores. Additionally, there was no significant interaction effect of hemisphere with AQ. However, results indicated a significant three-way interaction of emotion, hemisphere, and AQ, a significant main effect of AQ on N170 amplitude, and no significant main effect of BVAQ scores on N170 amplitude. Correlational analyses revealed a significant negative correlation for AQ scores and right hemisphere N170 amplitude for fearful and neutral faces, as well as left hemisphere N170 amplitude in response to neutral faces. Meaning, that as AQ scores increased, right hemisphere N170 amplitudes became more negative for fearful and neutral faces, and that left hemisphere N170 amplitudes only became more negative for neutral faces. Regarding N170 latency, results showed no significant effects or interactions, suggesting that the degree of alexithymia or autism severity, emotional category, or hemisphere did not significantly influence the efficiency of structural encoding of faces. Results surrounding N250 response amplitude did not indicate any significant main effects or interactions for either emotional category or AQ and BVAQ scores. However, results did show a significant main effect of BVAQ on N250 latency, suggesting that alexithymia significantly influenced higher-order emotion processing, not autism. Additionally, correlational analyses found a significant negative correlation between BVAQ scores and N250 latency for neutral faces. Meaning that N250 responses occurred earlier as the severity of alexithymia as indexed by the BVAQ increased. The authors suggest that this efficiency in processing neutral faces found in alexithymia may be indicative of faster processing of facial information that is not related to emotional content.

Together, these results would suggest that there are potential differences in ERP responses related to face perception for ASD and alexithymic individuals. Specifically, the N170 for autistic individuals and N250 for alexithymic individuals. However, it should be noted that the observation of a more negative N170 observed in ASD individuals ran contrary to the hypothesis of Desai et al. (2019). The authors suggest that this may be due to post hoc analyses revealing that the negative correlation was largely driven by the AQ subscale of Attention Switching, which may imply that a reduced ability to switch attention may have caused the larger N170 amplitudes for fearful and neutral faces. Additionally, the authors suggesting that neutral faces facilitate an earlier N250 latency observed in alexithymic individuals may be driven by a lack of emotional information present in neutral expression would require further examination into any potential differences in N250 responses for other emotional expressions (e.g., neutral compared to anger, neutral compared to happy, etc.). Finally, another limitation of this study was only examining autism and alexithymia severity scores in NT individuals. Future work should aim to examine the influence of clinical diagnoses of autism and/or alexithymia on ERPs related to facial expression processing.

CHAPTER 6. THE CURRENT STUDY

It has been proposed that alexithymic individuals possess a trait that makes it difficult for them to verbally describe internal emotional states and the emotional states of others. And according to the alexithymia hypothesis, the FER deficits observed in ASD may be better attributed to frequently co-occurring alexithymia (e.g., Bird & Cook, 2013). However, despite these propositions, almost all previous research has used lexical labels to categorize facial expressions when investigating the recognition abilities of ASD and NT individuals with varying levels of alexithymia. Using lexical labels to categorize or verbalize which emotional category an expression belongs to makes it difficult to tell whether reduced FER abilities in ASD may be better explained by comorbid alexithymia.

These gaps in face-related research raise the following question: Are the observed reduced FER abilities in ASD and alexithymia caused by atypical perceptual processes of facial expressions or by abnormal recognition processes involving verbal labeling of facial expressions? This question was addressed by asking 59 participants with ASD or who are NT, with varying levels of alexithymia, to partake in an eye-tracking study as they completed one of two facial expression recognition-related tasks (i.e., a lexical labeling of facial expressions task and a pictorial facial expression matching task). Eye tracking data was collected to better understand any potential group differences in eye movement behaviors (i.e., AOI fixation count and dwell times) for facial expression processing, as well as to reveal any potential underlying group differences for features of eye movement behaviors during facial expression processing examined via machine learning methods. ASD and NT individuals had their ASD and alexithymia severity scores calculated via the AQ-10 and TAS-20, respectively. Our working first hypothesis was that deficits in facial expression recognition often reported in ASD would be better explained by comorbid alexithymia, the alexithymia hypothesis.

Based on this hypothesis, we first predicted that ASD individuals would exhibit differential eye movement patterns during facial expression processing compared to NT individuals, such that ASD individuals would spend less time attending to internal facial features than NT individuals, consistent with previous studies (e.g., Kang et al., 2020). Importantly, we expected that there could be differential eye movement patterns between ASD individuals with higher alexithymia-severity scores compared to ASD individuals with lower alexithymia-severity scores. We predicted that eye movement patterns of high-alexithymic ASD individuals would be similar to high-alexithymic NT individuals and that low-alexithymic ASD individuals would have eye movement patterns comparable to low-alexithymic NT individuals. Regarding gaze patterns, previous research has indicated that ASD and alexithymic individuals spend less time attending to the eyes than NT individuals but that ASD and alexithymic individuals do not significantly differ from one another in their eye: mouth gaze ratios (Bird et al., 2011). However, other results have indicated that during the presentation of dynamic facial expressions, alexithymia was a better predictor of reduced eye gaze than autism (Cuve et al., 2021).

Our second working hypothesis was that alexithymia is caused by abnormal recognition processes involving verbal labeling of facial expressions rather than by atypical perceptual processing of facial expressions (Lane et al., 1996). If alexithymia is a consequence of impaired symbolic representation of emotion and if FER ability is measured by using stimuli or responses that involve an emotional word, then poor performance may be linked to an inability to comprehend and use the emotional terms appropriately. Based on this hypothesis, we predicted that we would find a difference in performance between our two tasks, wherein individuals with higher levels of alexithymia would perform worse than individuals with lower levels of alexithymia at FER when tasked with using lexical labels to categorize expressions. Finally, we predicted that individuals with higher levels of alexithymia would display increased FER performance when tasked with matching facial expressions compared to lexically labeling them. These findings would be able to provide insight into differences in facial expression processing amongst different diagnostic and NT groups, and through machine learning, may be able to provide an alternative diagnostic tool for ASD and alexithymia.

6.1. Materials and Methods

6.1.1. Participants

59 participants comprised of 10 ASD (4 High-A and 6 Low-A individuals) and 49 neurotypical individuals (20 High-A and 29 Low-A individuals) were recruited from Florida Atlantic University's (FAU) SONA system. Participants were compensated with course credit reflective of the value of participation in the approximately two-hour experiment. Inclusion criteria required all participants to have normal or corrected-to-normal vision. If a participant had corrected-to-normal vision, they were asked to wear contacts to ensure the proper performance of our eye tracker.

6.1.2. Testing Environment

Participants were seated in a darkened room, painted with black light-absorbing paint. In the room, participants sat in front of a display monitor and an EyeLink 1000 Plus Desktop Mount with an eye-to-monitor distance set to 1.75x the display width of our display monitory (i.e., 104 cm away from a 59.5-cm wide display monitor). Distance from the monitor and head position was fixed for the experiment via a chinrest. All stimuli in this study were created and using Adobe Photoshop 2020, MATLAB, and Psychtoolbox-3.

6.1.3. Stimuli & Design

Regarding the recordings of facial expressions, participants were randomly assigned to one of two FER tasks (Lexical Labeling of Expressions task and Facial Expression Matching task), where five expressions (i.e., angry, happy, fear, neutral, and sad) were presented to each participant. For the presentation of realistic expressions, stimuli were derived from 20 different actors (10 males, 10 females) taken from the Karolinska Directed Emotional Faces (KDEF) database (Lundqvist et al., 1998). Furthermore, images were manipulated for intensity by being morphed via FantaMorph to remove any potential ceiling effects (version 5.4.2 Deluxe, Abrosoft, Beijing, China, http://www.fantamorph.com/). Each emotional expression of each actor was morphed along 4 frames (i.e., low – 40%, medium – 60%, high – 80% full intensity – 100%) created by increasing the intensity of the emotional expression surrounding two images. The first image was a neutral image, and the second was an emotional image (either angry or happy). For the presentation of schematic facial expressions, schematic stimuli were consistent with those seen in Sawada and Sato (2015). The selection process for the

KDEF images required that the actor within the picture be facing the camera from a straight angle with a frontal gaze to collect all the necessary facial components for facial expression discrimination. All the images were adjusted to black & white by using the default black & white setting found within Adobe Photoshop 2020 and had a stimulus size in visual angle equal to 7.26°.

A group of participants was assigned to the Lexical Labeling of Expressions (LLE) task, where participants were asked to attend to a centrally located target face. Then via button-press, indicate via mouse click which expression label best describes the target face, followed by indicating via mouse click their perceived-valence rating of the target face (i.e., a Likert-scale rating, ranging from 1 to 7, where a rating of 1 was indicative of an extremely negative valence rating and 7 is indicative of an extremely positive valence rating). The order of the task was as such. First, a fixation cross was presented for 500 ms, followed by a random target expression (of either neutral, or low, medium, or full intensity angry, happy, fear, or sad) for 3000 ms - consistent with other studies (e.g., Dalton et al., 2005) - then the lexical FER screen until a response was made, followed by the perceived valence rating screen that was presented until a response was made, and finally an intertrial interval (ITI) of 500 ms.

Another group of participants was assigned to the Facial Expression Matching (FEM) task followed the same procedure as the lexical labeling of facial expressions task. However, instead of describing which lexical label best describes the expression of the target face (e.g., 1 = angry), participants were instead asked to indicate which image (of a schematic facial expression) best matched the target face. The use of schematic facial expressions vs. realistic facial expressions allowed us to maintain consistent features across potential responses, as well as control for any race-influenced effects of response (Bi et al., 2022). Before the beginning of the experiment, participants participated in a familiarization phase where they were presented with 20 images derived from presentations of each of the expression stimuli and asked to complete the familiarization task twice (i.e., shown 40 images in total). A short break was given to the participants at the halfway point of the experiment and at any point if needed.

Upon entering the experiment room, participants first completed a questionnaire consisting of demographic information (i.e., gender, race, ethnicity, handedness), followed by the AQ-10, and TAS-20 measurements. Participants were labelled as ASD if their AQ-10 score was 6 or higher, and NT if their AQ-10 score was equal to or less than 5. Participants were labelled as High-A if their TAS-20 scores were 52 or higher, and Low-A if their TAS-20 scores were equal to or less than 51. Following the completion of the questionnaire, participants completed the WAIS-IV Matrix Reasoning task (age information was collected here as part of this task). After finishing the WAIS-IV Matrix Reasoning task, participants were then randomly assigned to either the LLE or FEM task, and provided with instructions on how to properly complete the task they were assigned. Next, participant partook in two practice blocks (i.e., the familiarization phase). Upon completion of the practice blocks, participants were exposed to 10 blocks of their randomly assigned task and were afforded the opportunity to take a short break at the halfway point of the experiment or at any point if needed.

6.1.4. Eye-Tracking Instrumentation, Configuration, and Processing

Eye signal recordings were obtained using an EyeLink 1000 Plus Desktop Mount. The head position was fixed via a chin rest, and participants were seated roughly 104 cm away from the monitor. The Desktop Mount was set 50 – 55 cm from the front of the chin rest. Eye-related signals were continually recorded at a sampling rate of 1000 Hz. Eye movement data was analyzed using MATLAB. Eight areas of interest (AOI) were selected for the analysis of target photos. These AOIs consisted of screen, image area, face, screen background (screen – image area AOI), image area no face (image area – face AOI), eyes, right eye, left eye, mouth, and nose. To avoid counting unconscious gazes, a 60 ms threshold will be applied.

6.1.5. Decoding Analysis Overview

We trained a Convolutional Neural Network (CNN) to classify patterns of eye movements during facial expression processing between ASD and NT individuals with high (High-A) and low alexithymic (Low-A) traits. We used a CNN for its advantages in high performance for image classification and the ability to remain highly accurate even when using a limited training set (Klaib et al., 2021). Machine learning methods for ASD and NT classification have been implemented in the past, such as by Kang et al. (2020) when classifying ASD and NT children when presented with neutral faces of own-race or other-race stimuli. Cross-validation methods were implemented to ensure that the classifier was performing correctly. Cross-validation was performed 10 times with random split data set (a two-thirds train to one-third test), consistent with Bae and Luck (2018) and Smith and Smith (2019). The data for decoding analyses were comprised of single-trial scanpath information image plots stemming from both the LLE and FEM Tasks with an equal number of inputs into the classifier per diagnostic group. Decoding was only considered correct if the classifier accurately determined which group an individual belonged to.

6.1.5.1. Normalize the Data (Preprocess). The following preprocessing procedures were implemented to ensure only high-quality, relevant scanpath information were fed into the classifier; this is illustrated in Figure 1. First, fixations outside the presentation screen's dimensions (i.e., 2560 x 1440 pixels) were not plotted in the scanpath plots. Secondly, any blink information was also discarded from plotting. Lastly, if fewer than three fixations occurred within a given trial, then that trial was discarded from plotting.

6.1.5.2. Plotting the Scanpaths. After preprocessing, the trials which were not discarded had their individual scanpaths plotted. Scanpath plots were created via fixation reports from both the LLE and the FEM tasks. For these plots, a circle was plotted at the location of each fixation, and a line was plotted for each saccade or transition between (x_t, y_t) to (x_{t+1}, y_{t+1}), where t represents time points within the trial window and x and y represent pixel coordinates of the presentation window. These plots were then plotted onto an area representative of the original presentation window dimensions to best capture the looking behavior of participants during the experiment. Plots were created with a mirrored y-axis to match the origin locations of the presentation screen. Nine unique plots were created per trial, first a plot that contained the scanpath from the right eye only, followed by a plot that included the scanpath from the left eye only, and a plot with the scanpath from both eyes. These three distinct eye plots were then manipulated to display spatial information, spatial-temporal information, and spatial-temporal-ordinal information. The spatial information plots were created via the saccadic information described earlier. The spatial-temporal information utilized the same technique for plotting saccadic information, but additionally the circular markers representing location of fixation were scaled in size to represent fixation duration. Finally, the spatial-temporal-ordinal plots followed the same steps as the spatial-temporal plots, but additionally added a numerical counter to each fixation location that represents the order of fixation (e.g., 1st fixation being marked with a "1", 2nd fixation being marked with a "2", etc.). With the above scanpath plotting procedures there would be, at maximum, 2,880 scanpath plots/participant (4 emotional faces x 4 expression intensities x 2 genders x 10 blocks x 9 unique scanpath plots) before any trial removal. This would equate to 169,920 scanpath plots prior to any trial removal (2,880 images/ participant x 59 participants).

These plots were then be transformed into images and further preprocessed to allow for greater classification performance (e.g., Wang et al., 2017) by being trimmed to contain only the relevant scanpath information (i.e., removing any negative white space in which fixations did not occur). This was accomplished by calculating the minimum and maximum locations of the plotted pixels for the scanpath, then cropping to remove any empty areas. After trimming, these plots were transformed into PNG images to be later fed into the classifier. Lastly, images were padded and scaled to 224x224 pixels, thus reducing dimensionality confounds by decreasing the number of features under consideration (Cilia et al., 2021). These image preparation steps are illustrated in Figure 2, and an example of the transition between original and transformed scanpath images can be seen in Figure 3.

6.1.5.3. Augmenting the Data. In order to increase the number of inputs to the classifier, data augmentation procedures were implemented for our ASD groups. Data augmentation has been shown to be a reliable technique for improving classifier performance (Xu et al., 2016; Wang et al., 2017; Ahmed et al., 2022). Specifically, we used a form of lossless transformation known as mirroring, in which each of the images of the scanpaths were flipped along their horizontal axis (left-right mirroring) to create what appears to the classifier as a new, entirely unique image but still retains the information from the original images. Figure 4 highlights this form of data augmentation by showing an original image of a raccoon and a mirrored image of a raccoon. If both images were fed to a classifier, they would be unique inputs but still contain the same underlying information and meaning. This form of transformation teaches the model to ignore the precise image presentation and instead focus on the unchanging underlying information. Using this form of augmentation enabled us to end up with roughly 190K images to be used in training and testing.

6.1.6. Statistical Analysis Overview

Concerning behavioral performance, we used the unbiased hit rate (Wagner, 1993) as our measurement for facial expression recognition (FER) accuracy across both tasks. Using the unbiased hit rate allowed us to account for any behavioral response tendencies of the participants. For instance, independent of expression presentation, a participant may have a bias towards responding with "happy" as the accurate emotional expression that best describes this stimulus presentation. In this example, the participants raw hit rate for happy expressions would be inflated. The unbiased hit rate accounts for biases in category judgment paradigms and is calculated as: $Hu = \left(\frac{Ai}{Bi}\right) * \left(\frac{Ai}{Ci}\right)$, where "Ai" corresponds to the number of hits, "Bi" corresponds to the number of trials where "i" is the target, and "Ci" corresponds to the frequency of "i" responses for both hits and false alarms.

To examine our prediction that High-A individuals will display increased FER performance for all expression intensities during the Facial Expression Matching (FEM) task compared to the Lexical Labeling of Expressions (LLE) task, a 2x4x4 mixed model ANOVA was conducted for High-A individuals. Our between-subject factor was task (the LLE task, the FEM task) and our within-subjects factors were expression (angry, happy, fearful, sad) and expression intensity (40%, 60%, 80%, 100%). To compare the FER performance between ASD and NT groups for all expression intensities within the LLE Task or within the FEM task, two separate 2x4x4 mixed model ANOVAs for each task were conducted. Our between-subjects factors were ASD group (ASD, NT) or Alexithymia group (High-A, Low-A), and our within-subjects factors were expression (angry, happy, fearful, sad) and expression intensity (40%, 60%, 80%, 100%). To examine our predictions that there will be differential eye movement patterns between ASD and NT individuals as well as High-A and Low-A individuals, comparisons between groups and image AOI fixation and dwell times were performed via paired-samples t-tests.

Additionally, to examine if there was a potential restriction of valence ratings for ASD individuals consistent with prior research (e.g., Ben-Shalom et al., 2006), as well as a potential restriction of valence for alexithymic individuals, another 2x4x4 Mixed Model ANOVA was conducted. Our between-subjects factors were ASD group (ASD, NT) or Alexithymia group (High-A, Low-A), and our within-subjects factors were valence ratings expression (angry, happy, fearful, sad) and expression intensity (40%, 60%, 80%, 100%). Finally, correlation analyses were conducted to examine if a participant's AQ-10 score was associated with their TAS-20 score, and if their AQ-10 or TAS-10 scores were associated with their WAIS-IV Matrix Reasoning scores.

6.2. RESULTS

6.2.1. Demographic Comparisons

Regarding demographic comparisons between ASD and NT individuals, there was no significant difference (p = .16) in age between ASD (M = 20.46, SD = 2.38) and NT (M = 19.50, SD = 1.93). We did, however, find a significant difference (p < .001) in AQ-10 scores between ASD and NT individuals, with ASD (M = 7.00, SD = 1.10; *Range* = 6 – 9) individuals having significantly higher AQ-10 scores than NT (M = 3.02, SD = 1.27; *Range* = 0 – 5) individuals. Additionally, a significant difference (p = .05) in TAS-20 scores between ASD and NT individuals was found, with ASD (M = 54.55, SD = 10.37; *Range* = 40 – 68) individuals having significantly higher TAS-20 scores than NT (M = 49.22, SD = 7.70; *Range* = 37 – 72) individuals. Furthermore, a significant difference was also found in WAIS-IV Matrix Reasoning scores between ASD and NT individuals, with ASD (M = 15.46, SD = 4.13; *Range* = 8 – 22) having significantly worse Matrix Reasoning scores than NT (M = 19.04, SD = 3.88; *Range* = 7 – 24).

Regarding demographic comparisons between High-A and Low-A individuals, no significant difference (p = .39) in age between High-A (M = 20.46, SD = 2.38) and Low-A (M = 19.50, SD = 1.93) individuals was found. Furthermore, no significant difference (p < .001) in AQ-10 scores between High-A and Low-A individuals was found. However, a significant difference (p < .001) in TAS-20 scores between High-A and Low-A individuals was found, with High-A (M = 58.11, SD = 5.52) individuals having significantly higher TAS-20 scores than Low-A (*M* = 44.45, *SD* = 4.46) individuals. Finally, no significant difference in WAIS-IV Matrix Reasoning scores between High-A and Low-A individuals was found.

6.2.2. Correlational Analyses

Correlational results suggest a negative correlation between AQ-10 scores and WAIS-IV Matrix Reasoning scores (*rs* = -0.28, *p* = 0.03), suggesting a potential negative relationship between autism severity and our IQ measure (the higher autism severity, the lower IQ score). No other correlations were shown to be significant in our analysis. Graphs of the above-mentioned correlations are shown in Figure 5.

6.2.3. Differences in Eye Movement Patterns

Regarding eye movement patterns, our predictions were that there would be, 1) differential eye movement patterns between ASD and NT individuals, 2) differential eye movement patterns between High-A and Low-A individuals, 3) similar eye movement patterns for ASD High-A individuals and NT High-A individuals, and 4) similar eye movement patterns for ASD Low-A individuals and NT Low-A individuals. To examine these predictions paired-samples t-tests for number of AOI fixations and dwell times were performed. Regarding comparisons between High-A and Low-A individuals a significant difference for the number of fixations made to the nose region was found [$t(57) = 2.489$, $p = .015$], suggesting that High-A individuals (*M* = 1.865, *SD* = 0.918) made more fixations to the nose than Low-A individuals (*M* = 1.266, *SD* = 0.846). No other comparisons for AOI fixations or dwell times were significant for comparisons High-A and Low-A individuals. Comparisons between ASD and NT individuals revealed no significant differences for number of AOI fixations made or dwell times, suggesting similar eye movement behaviors for facial expression processing for ASD and NT individuals

Comparisons between ASD Low-A and NT Low-A individuals revealed significant differences between the number of fixations made to the screen background [$t(57) = 2.145$, $p = .035$] and the dwell times on the screen background [$t(57) = 2.606$, $p = .011$], suggesting that ASD Low-A individuals ($M = 0.016$, $SD = 0.846$) made more fixations to the screen background than NT Low-A individuals ($M = 0.003$, $SD = 0.013$), and that ASD Low-A individuals ($M = 2.641$, $SD = 4.890$) spent more time attending to the screen background than NT Low-A individuals ($M = 0.442$, $SD = 1.639$). However, due to the low fixation and dwell time values for the ASD Low-A and NT Low-A individuals, these results may be more reflective of unconscious gaze processes rather than facial expression processing. No other significant differences were found for comparisons regarding the number of AOI fixations made or dwell times between ASD Low-A and NT Low-A individuals. Additionally, when comparing between ASD High-A and NT High-A individuals, no significant differences for number of AOI fixations made or dwell times were found. These alexithymic-specific results when considering that differences between ASD Low-A and NT Low-A individuals may be more due to unconscious gaze processes, may suggest that during facial expression processing, individuals exhibit similar eye movement behaviors when matched on alexithymia severity.

6.2.4. Behavioral Results – Task Influence on FER Performance

To examine our prediction that High-A individuals will display increased FER performance during the FEM task compared to the LLE task, a 2x4x4 mixed model

ANOVA was conducted for High-A individuals. Our between-subjects factor was task (the FEM task, the LLE task) and our within-subjects factors were expression (angry, happy, fearful, sad) and expression intensity (40%, 60%, 80%, 100%). When comparing the FER performances for High-A individuals between the two tasks, analyses of unbiased hit rate showed significant interactions between emotion and task [$F(3, 72) = 5.907$, $p < .001$], and emotion and intensity [$F(9, 216) = 2.737$, $p = .005$], as well as main effects of emotion [$F(3, 72) = 5.907$, $p = .001$] and of intensity [$F(3, 72) = 176.800$, $p < .001$]. No other interactions were significant. Figure 6(a) illustrates that contrary to our predictions, it was also found that High-A individuals completing the FEM task ($M = .622$, $SE = .032$) performed significantly worse than High-A individuals completing the LLE task ($M = .756$, $SE = .035$) [$F(1, 24) = 7.857$, $p = .010$].

Bonferroni post-hoc tests revealed that High-A individuals assigned to the FEM task had significantly worse accuracy scores than High-A individuals assigned to the LLE task for presentations of angry ($M_{FEM} = .673$, $SE_{FEM} = .036$; $M_{LLE} = .859$, $SE_{LLE} = .039$) ($p = .002$) and sad ($M_{FEM} = .554$, $SE_{FEM} = .040$; $M_{LLE} = .785$, $SE_{LLE} = .043$) ($p < .001$) facial expressions, but similar accuracy scores for fearful ($M_{FEM} = .909$, $SE_{FEM} = .023$; $M_{LLE} = .936$, $SE_{LLE} = .025$) ($p = .419$) and happy ($M_{FEM} = .700$, $SE_{FEM} = .032$; $M_{LLE} = .786$, $SE_{LLE} = .034$) ($p = .078$) facial expression presentations. These results suggest that High-A individuals may have a greater understanding of the emotional words or lexical labels representative of angry and sad than schematic representations of sad. Additionally, similar accuracies across both tasks for happy and fearful expressions suggest that schematic representations of emotions may be sufficient to allow for similar accuracy for High-A individuals across both tasks. Additionally, the comparable accuracy scores across both tasks for happy may be reflective of a positive classification advantage (Liu et al., 2013), while similar accuracies for fear may be representative of a threat detection advantage for fear (Hedger et al., 2015).

Bonferroni post-hoc tests also revealed that regarding accuracy for intensity-specific angry presentations, accuracy scores for 40% angry (M = .495, SE = .035) were significantly lower than 60% (M = .782, SE = .038), 80% (M = .877, SE = .027), and 100% (M = .911, SE = .027) intensities of angry ($p < .001$, respectively), as well as 60% angry being significantly lower than 80% ($p = .002$) and 100% angry ($p < .001$), but that 80% angry did not significantly differ from angry 100% presentations ($p = .206$). Bonferroni post-hoc tests also showed that regarding accuracy for intensity-specific fear presentations, accuracy scores for 40% fear (M = .744, SE = .048) were significantly lower than 60% (M = .965, SE = .018), 80% (M = .987, SE = .006), and 100% (M = .994, SE = .003) intensities of fear ($p < .001$, respectively), but that 60% fear did not significantly vary from 80% ($p = .806$) and 100% fear ($p = .489$), and that 80% fear did not significantly differ from fear 100% presentations ($p = .276$). Regarding accuracy for intensity-specific presentations of happy, Bonferroni post-hoc revealed that accuracy scores for 40% happy (M = .465, SE = .033) were significantly lower than 60% (M = .784, SE = .030), 80% (M = .837, SE = .024), and 100% (M = .887, SE = .022) intensities of happy ($p < .001$, respectively), and that accuracy scores for 60% happy were significantly lower than 80% ($p = .007$) and 100% happy ($p < .001$), and that accuracy scores for 80% happy were significantly lower than happy 100% presentations ($p = .041$). Finally, Bonferroni post-hoc tests regarding intensity-specific presentations of sad revealed that accuracy scores for 40% sad (M = .439, SE = .029) were significantly lower than 60% ($M = .685$, $SE = .039$), 80% ($M = .780$, $SE = .034$), and 100% ($M = .774$, $SE = .030$) intensities of sad ($p < .001$, respectively), and that accuracy scores for 60% sad were significantly lower than 80% ($p = .001$) and 100% sad ($p = .002$), and that accuracy scores for 80% sad were not significantly different from sad 100% presentations ($p = 1.000$). Together these results would suggest that individuals perform less accurately when categorizing lower intensities of expression than for higher intensities of expressions, consistent with findings from previous research (e.g., Griffiths et al., 2019). Figure 7(a) illustrates the significant emotion by task interaction, highlighting the differences in FER performance for High-A individuals based on emotional expression category, showing that High-A individuals in the FEM task performed worse for categorizing all expressions than High-A individuals in the LLE task. Figure 7(b) illustrates the differences in FER performance for High-A individuals based on intensity of expression category, showing that High-A individuals in the FEM task performed worse at all expression intensities compared to High-A individuals in the LLE task, although this interaction was found to not be significant.

To examine if our above results were potentially a product of task difficulty, another 2x4x4 mixed model ANOVA was conducted with the same between-subject and within-subjects factors, but instead ran on Low-A individuals. When comparing the FER performances for Low-A individuals between the two tasks, analyses of unbiased hit rate again showed significant interactions between emotion and task [$F(3, 93) = 4.185$, $p = .008$], and emotion and intensity [$F(9, 279) = 5.654$, $p < .001$], as well as main effects of emotion [$F(3, 93) = 70.727$, $p < .001$] and of intensity [$F(3, 93) = 193.618$, $p < .001$]. No other interactions were significant. Figure 7(b) shows that we again found when individuals completed the FEM task ($M = .614$, $SE = .027$) their FER performance was significantly worse than individuals that completed the LLE task ($M = .754$, $SE = .029$) [$F(1, 24) = 7.857$, $p = .001$].

Bonferroni post-hoc tests revealed that Low-A individuals assigned to the FEM task had significantly worse accuracy scores than Low-A individuals assigned to the LLE task for presentations of angry ($M_{FEM} = .554$, $SE_{FEM} = .042$; $M_{LLE} = .752$, $SE_{LLE} = .046$) ($p = .003$), happy ($M_{FEM} = .573$, $SE_{FEM} = .037$; $M_{LLE} = .717$, $SE_{LLE} = .041$) ($p = .014$), and sad ($M_{FEM} = .447$, $SE_{FEM} = .034$; $M_{LLE} = .632$, $SE_{LLE} = .037$) ($p < .001$) facial expressions, but similar accuracy scores for fearful ($M_{FEM} = .880$, $SE_{FEM} = .019$; $M_{LLE} = .916$, $SE_{LLE} = .020$) ($p = .209$) facial expression presentations. These results suggest that Low-A individuals perform similarly for categorization of fear faces regardless of task due to a potential threat detection advantage of fear (e.g., Hedger et al., 2015). Bonferroni post-hoc tests also revealed that regarding accuracy for intensity-specific angry presentations, accuracy scores for 40% angry ($M = .427$, $SE = .037$) were significantly lower than 60% ($M = .674$, $SE = .035$), 80% ($M = .763$, $SE = .032$), and 100% ($M = .747$, $SE = .034$) intensities of angry ($p < .001$, respectively), as well as 60% angry being significantly lower than 80% ($p < .001$) and 100% angry ($p = .002$), but that 80% angry did not significantly differ from angry 100% presentations ($p = 1.00$). Bonferroni post-hoc tests also showed that regarding accuracy for intensity-specific fear presentations, accuracy scores for 40% fear ($M = .743$, $SE = .034$) were significantly lower than 60% ($M = .920$, $SE = .018$), 80% ($M = .951$, $SE = .009$), and 100% ($M = .978$, $SE = .007$) intensities of fear ($p < .001$, respectively), that 60% fear did not significantly vary from 80% ($p = .542$) but did from 100% fear ($p = .007$), and that 80% fear did not significantly differ from fear 100%

presentations (p = .114). Regarding accuracy for intensity-specific presentations of happy, Bonferroni post-hoc revealed that accuracy scores for 40% happy (M = .405, SE = .027) were significantly lower than 60% (M = .671, SE = .034), 80% (M = .751, SE = .031), and 100% (M = .753, SE = .033) intensities of happy (p < .001, respectively), and that accuracy scores for 60% happy were significantly lower than 80% (p = .010) and 100% happy (p = .016), and that accuracy scores for 80% happy did not significantly vary from happy 100% presentations (p = 1.00). Finally, Bonferroni post-hoc tests regarding intensity-specific presentations of sad revealed that accuracy scores for 40% sad (M = .280, SE = .026) were significantly lower than 60% (M = .528, SE = .031), 80% (M = .653, SE = .028), and 100% (M = .695, SE = .035) intensities of sad (p < .001, respectively), and that accuracy scores for 60% sad were significantly lower than 80% and 100% sad (p < .001, respectively), and that accuracy scores for 80% sad were not significantly different from sad 100% presentations (p = .384).

Coupled with the above High-A results, these findings may also suggest that the decrease in FER performance for the FEM matching task for all individuals may due to task difficulty, contrary to our prediction of expecting an increase in performance for the FEM task when controlling for alexithymia severity. Figure 8(a) illustrates the significant emotion by task interaction, highlighting the differences in FER performance for Low-A individuals based on emotional expression category, showing that Low-A individuals in the FEM task performed worse for categorizing all expressions than Low-A individuals in the LLE task. Figure 8(b) illustrates the differences in FER performance for Low-A individuals based on intensity of expression category, showing that Low-A individuals in the FEM task performed worse at all expression intensities compared to Low-A individuals in the LLE task, although this interaction was not significant.

6.2.5. Behavioral Results – Lexical Labeling of Expressions (LLE) Task

To compare groups on their FER performance for all expression intensities during the LLE Task, two separate 2x4x4 mixed model ANOVAs were conducted, one for each diagnostic group. Our between-subjects factors were ASD group (ASD, NT) or Alexithymia group (High-A, Low-A), and our within-subjects factors were expression (angry, happy, fearful, sad) and expression intensity (40%, 60%, 80%, 100%). When comparing between ASD and NT individuals, analyses of unbiased hit rate for the LLE task revealed significant interactions between intensity and group [$F(3,75) = 5.184$, $p = .003$], as well as for emotion and intensity [$F(9, 225) = 3.420$, $p < .001$]. Additionally, results showed a main effect of emotion [$F(3, 75) = 26.986$, $p <.001$], as well as a main effect of intensity [$F(3, 75) = 148.244$, $p <.001$]. No other main effects were significant. No other interactions or main effects were significant.

Bonferroni post-hoc tests revealed that during the LLE task ASD individuals had significantly worse accuracy scores at 40% ($M = .461$, $SE = .045$) intensities of expression compared to 60% ($M = .838$, $SE = .046$), 80% ($M = .877$, $SE = .035$), and 100% intensities ($M = .879$, $SE = .038$) ($p < .001$, respectively). These same patterns of results were also revealed for NT individuals where they performed significantly worse at 40% ($M = .544$, $SE = .019$) intensities of expressions compared to 60% ($M = .775$, $SE = .019$), 80% ($M = .836$, $SE = .015$), and 100% intensities ($M = .858$, $SE = .016$), as well as lower accuracy scores for 60% intensities compared to 100% intensities ($p <.001$, respectively). These results confirm that individuals perform worse at recognizing lower intensities of expression. Bonferroni post-hoc tests revealed that regarding accuracy for intensity-specific angry presentations, accuracy scores for 40% angry (*M* = .526, *SE* = .045) were significantly lower than 60% (*M* = .803, *SE* = .038), 80% (*M* = .846, *SE* = .027), and 100% (*M* = .843, *SE* = .033) intensities of angry ($p < .001$, respectively), but that 60% angry was not significantly different than 80% ($p = .926$) or 100% angry ($p = 1.00$), and that 80% angry did not significantly differ from angry 100% presentations ($p = 1.00$). Bonferroni post-hoc tests also showed that regarding accuracy for intensity-specific fear presentations, accuracy scores for 40% fear (*M* = .704, *SE* = .043) were significantly lower than 60% (*M* = .803, *SE* = .038), 80% (*M* = .846, *SE* = .027), and 100% (*M* = .843, *SE* = .033) intensities of fear ($p < .001$, respectively), that 60% fear did not significantly vary from 80% ($p = .751$) or 100% fear ($p = .247$), and that 80% fear did not significantly differ from fear 100% presentations ($p = 1.00$). Regarding accuracy for intensity-specific presentations of happy, Bonferroni post-hoc revealed that accuracy scores for 40% happy (*M* = .453, *SE* = .038) were significantly lower than 60% (*M* = .768, *SE* = .035), 80% (*M* = .807, *SE* = .035), and 100% (*M* = .823, *SE* = .033) intensities of happy ($p < .001$, respectively), but that accuracy scores for 60% happy were not significantly different from 80% ($p = 1.000$) or 100% happy ($p = .689$), and that accuracy scores for 80% happy did not significantly vary from happy 100% presentations ($p = 1.00$). Finally, Bonferroni post-hoc tests regarding intensity-specific presentations of sad revealed that accuracy scores for 40% sad (*M* = .328, *SE* = .040) were significantly lower than 60% (*M* = .696, *SE* = .040), 80% (*M* = .786, *SE* = .031), and 100% (*M* = .825, *SE* = .033) intensities of sad ($p < .001$, respectively), and that accuracy scores for 60% sad were significantly lower than 80% ($p = .037$) and 100% sad ($p = .002$), and that accuracy scores for 80% sad were not significantly different from sad 100% presentations (p = .647) (Figure 9).

Analyses of unbiased hit rate for the LLE task for High-A compared to Low-A individuals showed a significant interaction between emotion and intensity [$F(9, 225) = 5.561$, $p < .001$], as well as a main effect of emotion [$F(3, 75) = 61.287$, $p < .001$], and a main effect of intensity [$F(3, 75) = 192.046$, $p < .001$]. No other interactions or main effects were significant. Bonferroni post-hoc tests revealed that during the LLE task, High-A and Low-A individuals had significantly worse accuracy scores at 40% angry (M = .539, SE = .032) compared to 60% (M = .776, SE = .027), 80% (M = .841, SE = .019), and 100% (M = .851, SE = .024) intensities of angry ($p < .001$, respectively), and that 60% angry was significantly different from 80% (p = .028) and 100% angry (p = .038), but that 80% angry did not significantly differ from angry 100% presentations (p = 1.00). Bonferroni post-hoc tests also showed that regarding accuracy for intensity-specific fear presentations, accuracy scores for 40% fear (M = .748, SE = .032) were significantly lower than 60% (M = .955, SE = .012), 80% (M = .978, SE = .006), and 100% (M = .980, SE = .007) intensities of fear ($p < .001$, respectively), that 60% fear did not significantly vary from 80% fear (p = .571) but did from 100% fear (p = .035), and that 80% fear did not significantly differ from fear 100% presentations (p = 1.00). Regarding accuracy for intensity-specific presentations of happy, Bonferroni post-hoc revealed that accuracy scores for 40% happy (M = .481, SE = .028) were significantly lower than 60% (M = .753, SE = .025), 80% (M = .791, SE = .024), and 100% (M = .815, SE = .024) intensities of happy ($p < .001$, respectively), but that accuracy scores for 60% happy were not significantly different from 80% (p = .855) or 100% happy (p = .093), and that accuracy scores for 80% happy did not significantly vary from happy 100% presentations (p = .592). Finally, Bonferroni post-hoc tests regarding intensity-specific presentations of sad revealed that accuracy scores for 40% sad (M = .360, SE = .029) were significantly lower than 60% (M = .658, SE = .029), 80% (M = .754, SE = .023), and 100% (M = .798, SE = .024) intensities of sad (p < .001, respectively), and that accuracy scores for 60% sad were significantly lower than 80% and 100% sad (p < .001, respectively), and that accuracy scores for 80% sad were not significantly different from sad 100% presentations (p = .093) (Figure 10).

6.2.6. Behavioral Results – Facial Expression Matching (FEM) Task

To compare groups on their FER performance for all expression intensities during the FEM Task, two separate 2x4x4 mixed model ANOVAs were conducted, one for each diagnostic group. Our between-subjects factors were ASD group (ASD, NT) or Alexithymia group (High-A, Low-A), and our within-subjects factors were expression (angry, happy, fearful, sad) and expression intensity (40%, 60%, 80%, 100%). When comparing between ASD and NT individuals, analyses of unbiased hit rate for the FEM task showed a main effect of emotion [F(3, 90) = 63.275, p <.001] and of intensity [F(3, 90) = 117.122, p <.001]. No other main effects or interactions were significant (Figure 11).

Analyses of unbiased hit rate for the FEM task for High-A and Low-A individuals showed an interaction between emotion and intensity [F(9, 270) = 3.985, p < .001], a main effect of emotion [F(3, 90) = 97.162, p <.001], and a main effect of intensity [F(3, 90) = 188.745, p <.001]. No other interactions or main effects were significant. Bonferroni post-hoc tests revealed that during the FEM task, High-A and Low-A individuals had significantly worse accuracy scores at 40% angry ($M = .370$, $SE = .043$) compared to 60% ($M = .690$, $SE = .042$), 80% ($M = .794$, $SE = .031$), and 100% ($M = .839$, $SE = .031$) intensities of angry ($p < .001$, respectively), and that 60% angry was significantly different from 80% and 100% angry ($p < .001$, respectively), but that 80% angry did not significantly differ from angry 100% presentations ($p = 1.00$). Bonferroni post-hoc tests also showed that regarding accuracy for intensity-specific fear presentations, accuracy scores for 40% fear ($M = .746$, $SE = .044$) was significantly lower than 60% ($M = .929$, $SE = .022$), 80% ($M = .967$, $SE = .008$), and 100% fear ($M = .989$, $SE = .004$) ($p < .001$, respectively), that 60% fear was not significantly different from 80% ($p = .198$) significantly different from 100% fear ($p = .028$), and that 80% fear did significantly differed from fear 100% presentations ($p = .014$). Regarding accuracy for intensity-specific presentations of happy, Bonferroni post-hoc revealed that accuracy scores for 40% happy ($M = .368$, $SE = .032$) were significantly lower than 60% ($M = .700$, $SE = .041$), 80% ($M = .810$, $SE = .034$), and 100% ($M = .844$, $SE = .033$) intensities of happy ($p < .001$, respectively), that accuracy scores for 60% happy were significantly lower than 80% and 100% happy ($p < .001$, respectively), and that accuracy scores for 80% happy did not significantly vary from happy 100% presentations ($p = .296$). Finally, Bonferroni post-hoc tests regarding intensity-specific presentations of sad revealed that accuracy scores for 40% sad ($M = .355$, $SE = .029$) were significantly lower than 60% ($M = .588$, $SE = .037$), 80% ($M = .697$, $SE = .033$), and 100% ($M = .683$, $SE = .038$) intensities of sad ($p < .001$, respectively), and that accuracy scores for 60% sad were significantly lower than 80% sad ($p < .001$) and 100% sad ($p = .040$) and that accuracy scores for 80% sad were not significantly different from sad 100% presentations (p = 1.00) (Figure 12).

6.2.7. Behavioral Results – Valence Ratings of Expression

To compare groups on their valence ratings for all expressions and expression intensities, two separate 2x4x4 mixed model ANOVAs were conducted, one for each diagnostic group. Our between-subjects factors were ASD group (ASD, NT) or Alexithymia group (High-A, Low-A), and our within-subjects factors were expression (angry, happy, fearful, sad) and expression intensity (40%, 60%, 80%, 100%). When comparing between ASD and NT individuals, analyses of valence ratings showed an intensity by group interaction [$F(3, 171) = 3.075$, $p = .029$], and main effects of emotion [$F(3, 171) = 406.449$, $p <.001$] and of intensity [$F(3, 171) = 25.657$, $p <.001$]. No other main effects or interactions were significant (Figure 11).

Bonferroni post-hoc tests showed that ASD and NT individuals marked lower intensities of expressions as more positive compared to higher intensities of expressions. It was revealed that ASD individuals rated 40% ($M = 4.220$, $SE = .135$) intensity of expressions as more positive than 60% ($M = 4.531$, $SE = .153$) ($p < .001$), 80% ($M = 4.625$, $SE = .167$) ($p < .001$), and 100% ($M = 4.700$, $SE = .181$) ($p = .001$) intensities of expression. These same patterns of results were also revealed for NT individuals who rated 40% ($M = 4.349$, $SE = .061$) intensity of expressions as more positive than 60% ($M = 4.487$, $SE = .069$), 80% ($M = 4.543$, $SE = .076$), and 100% ($M = 4.585$, $SE = .082$) intensities of expressions ($p < .001$, respectively). It was also shown that ASD individuals did not rate 60% intensity of expressions significantly different from 80% ($p = .530$) or 100% ($p = .234$) intensities of expression, nor did they rate 80% intensity of expressions significantly different from 100% (p = .355) intensities of expression. These same result patterns were also found for NT individuals, who did not rate 60% intensity of expressions significantly different from 80% (p = .152) or 100% (p = .051) intensity of expressions, nor did they rate 80% intensity of expression significantly different from 100% (p = .116) intensity of expressions. These results suggest that as individuals are presented with various intensities of expression, intensity influences the valence rating scores, such that lower intensities of expression are marked as more positive regardless of the expression presented, thus providing support for the argument that variations in emotional intensity can modulate emotional responses (e.g., Adolphs and Alpers (2010)).

When comparing between High-A and Low-A individuals, analyses of valence ratings showed an emotion by intensity interaction [$F(3, 225) = 5.561$, $p < .001$], and main effects of emotion [$F(3, 75) = 61.287$, $p <.001$] and of intensity [$F(3, 75) = 192.046$, $p <.001$]. No other main effects or interactions were significant (Figure 11). Bonferroni post-hoc tests showed that within expression category, individuals marked lower intensities of a particular expression as more positive compared to higher intensities of the same expression. It was revealed that individuals rated 40% (M = 4.877, SE = .068) angry as significantly more positive than 60% (M = 5.309, SE = .082), 80% (M = 5.570, SE = .099), and 100% (M = 5.739, SE = .1.08) (p < .001, respectively) intensities of angry. Additionally, regarding angry, it was shown that 60% intensities of angry were marked significantly more positive than 80% and 100% (p < .001, respectively) intensities of angry, and that 80% angry was rated as significantly more positive than 100% (p < .001) intensities of angry. It was also revealed that individuals rated 40% (M = 4.762, SE = .065) fear as significantly more positive than 60% (M = 5.185, SE = .089), 80% ($M = 5.377$, $SE = .101$), and 100% ($M = 5.548$, $SE = .113$) ($p < .001$, respectively) intensities of fear. It was also shown that individuals rated 60% intensities of fear as significantly more positive than 80% and 100% intensities of fear ($p < .001$), and that 80% intensities of fear were rated as significantly more positive than 100% intensities of fear ($p < .001$), respectively. Additionally, it was revealed that individuals rated 40% ($M = 3.049$, $SE = .074$) happy as significantly less positive than 60% ($M = 2.422$, $SE = .073$), 80% ($M = 1.988$, $SE = .056$), and 100% ($M = 1.669$, $SE = .054$) ($p < .001$, respectively) intensities of happy. Regarding happy, it was also shown that individuals rated 60% happy as significantly less positive than 80% and 100% intensities of happy ($p < .001$, respectively), and that 80% happy was rated as significantly less positive than 100% intensities of happy ($p < .001$, respectively) It was shown that individuals rated 40% ($M = 4.597$, $SE = .068$) sad as more positive than 60% ($M = 5.039$, $SE = .078$), 80% ($M = 5.279$, $SE = .089$), and 100% ($M = 5.444$, $SE = .095$) ($p < .001$, respectively) intensities of sad. Additionally, it was shown that regarding sad, 60% intensities of sad were rated as significantly more positive than 80% and 100% intensities of sad ($p < .001$, respectively), and that 80% intensities of sad were rated as significantly more positive than 100% intensities of sad ($p < .001$). Results of the emotion by intensity interaction would suggest that as individuals are presented with various intensities of a particular expression, intensity influences the valence rating scores, such that lower intensities of negative expressions (i.e., angry, fear, and sad) are marked as more positive than their higher intensity counterparts, and that higher intensities of positive expressions (i.e., happy) are rated as more positive than their lower intensity counterparts. Thus, again providing further support for the argument that intensity of emotional expressions can modulate emotional responses (e.g., Adolphs and Alpers (2010)).

6.2.8. Decoding Results – ASD High-A, ASD Low-A, NT High-A, and NT Low-A Individuals

To examine any potential emotion specific effects (ESE) regarding eye movement patterns, separate decoding analyses were performed for each emotion. To assess if the classifiers were performing significantly above chance level while accounting for sample size, a binomial cumulative distribution (Combrisson & Jerbi, 2015). Figure 14 shows the confusion matrices for classification accuracy and confusion percentages for all angry classification conditions for ASD High-A, ASD Low-A, NT High-A, and NT Low-A individuals. Figure 14(a) reveals that when spatial-temporal-ordinal scanpath images for angry were fed into the classifier, ASD Low-A individuals (51.75%) had the highest classification accuracy, followed by ASD High-A (47.08%), NT High-A (37.35%), and NT Low-A (36.96%). Figure 14(b) reveals that when spatial-temporal scanpath images for angry were fed into the classifier, ASD Low-A individuals (73.15%) had the highest classification accuracy, followed by ASD High-A (54.09%), NT High-A (35.02%), and NT Low-A (28.79%). Figure 14(c) reveals that when spatial scanpath images for angry were fed into the classifier, ASD Low-A individuals (71.21%) had the highest classification accuracy, followed by ASD High-A (38.91%), NT Low-A (34.24%), and NT High-A (29.96%).

Figure 15 shows the confusion matrices for classification accuracy and confusion percentages for all fear classification conditions for ASD High-A, ASD Low-A, NT High-A, and NT Low-A individuals. Figure 15(a) reveals that when spatial-temporal-ordinal scanpath images for fear were fed into the classifier, ASD High-A individuals (61.87%) had the highest classification accuracy, followed by ASD Low-A (54.47%), NT High-A (53.31%), and NT Low-A (47.08%). Figure 15(b) reveals that when spatial-temporal scanpath images for fear were fed into the classifier, ASD Low-A individuals (58.75%) had the highest classification accuracy, followed by ASD High-A (51.75%), NT Low-A (42.02%), and NT High-A (35.80%). Figure 15(c) reveals that when spatial scanpath images for fear were fed into the classifier, ASD Low-A individuals (60.31%) had the highest classification accuracy, followed by NT Low-A (44.36%), NT High-A (35.80%), and ASD High-A (33.85%).

Figure 16 shows the confusion matrices for classification accuracy and confusion percentages for all happy classification conditions for ASD High-A, ASD Low-A, NT High-A, and NT Low-A individuals. Figure 16(a) reveals that when spatial-temporal-ordinal scanpath images for happy were fed into the classifier, ASD High-A individuals and ASD Low-A individual (59.53%, respectively) had the highest classification accuracies, followed NT Low-A (52.53%), and NT High-A (49.81%). Figure 16(b) reveals that when spatial-temporal scanpath images for happy were fed into the classifier, ASD Low-A individuals (68.09%) had the highest classification accuracy, followed by NT High-A (58.75%), ASD High-A (49.03%), and NT Low-A (21.40%). Figure 16(c) reveals that when spatial scanpath images for happy were fed into the classifier, ASD Low-A individuals (57.98%) had the highest classification accuracy, followed by NT Low-A (48.64%), ASD High-A (48.25%), and NT High-A (36.96%).

Figure 17 shows the confusion matrices for classification accuracy and confusion percentages for all sad classification conditions for ASD High-A, ASD Low-A, NT High-

A, and NT Low-A individuals. Figure 17(a) reveals that when spatial-temporal-ordinal scanpath images for sad were fed into the classifier, ASD High-A individuals (69.65%) had the highest classification accuracy, followed by NT Low-A (66.54%), ASD High-A (47.86%), and NT High-A (38.13%). Figure 17(b) reveals that when spatial-temporal scanpath images for sad were fed into the classifier, ASD Low-A individuals (75.49%) had the highest classification accuracy, followed by NT High-A (40.08%), ASD High-A (36.58%), and NT Low-A (36.19%). Figure 17(c) reveals that when spatial scanpath images for sad were fed into the classifier, ASD Low-A individuals (70.04%) had the highest classification accuracy, followed by NT High-A (42.02%), ASD High-A (40.47%), and NT Low-A (35.41%).

Figure 18 shows the confusion matrices for classification accuracy and confusion percentages for all emotional expression classification conditions (i.e., angry, fear, happy, and sad scanpath images being fed into the classifier) for ASD High-A, ASD Low-A, NT High-A, and NT Low-A individuals. Figure 18(a) reveals that when spatial-temporal-ordinal scanpath images for all emotional expressions were fed into the classifier, ASD High-A individuals (75.65%) had the highest classification accuracy, followed by NT Low-A (58.29%), NT High-A (55.58%), and ASD High-A (48.21%). Figure 18(b) reveals that when spatial-temporal scanpath images for all emotional expressions were fed into the classifier, ASD Low-A individuals (71.10%) had the highest classification accuracy, followed by NT Low-A (61.20%), ASD High-A (55.00%), and NT High-A (32.01%). Figure 18(c) reveals that when spatial scanpath images for all emotional expressions were fed into the classifier, ASD Low-A individuals (80.70%) had the highest classification accuracy, followed by NT Low-A (48.50%), ASD High-A (47.24%), and NT High-A (32.78%).

Table 1 shows the decoding accuracy for all the classification conditions. Results indicate that all classification conditions performed significantly above chance level ($p <$ .001, respectively). Regarding ESE on overall decoding accuracy, eye movement patterns related to the processing of sad facial expressions yielded the highest accuracy compared to all other expressions. These results were observed when the classifier was fed with more relevant scanpath information being plotted onto the images (i.e., the Spatial-Temporal-Ordinal images). Results showed that for the Spatial-Temporal-Ordinal images, sad had the highest accuracy (55.55%), followed by happy (55.35%), fear (54.18%), and lastly angry (50.68%). Again, regarding scanpath information, results showed that when all expressions were fed into the classifier, the Spatial-Temporal-Ordinal classification yielded the highest decoding accuracy (59.43%), followed by Spatial-Temporal (54.83%), and Spatial (52.30%). This pattern of a decrease in classifier performance when less relevant scanpath information was fed to the classifier was also observed for all emotional expression conditions. For angry expression classification conditions, spatial-temporal-ordinal accuracy had the highest accuracy (50.68%), followed by spatial-temporal (47.76%), and spatial (43.58%). For fear expression classification conditions, spatial-temporal-ordinal had the highest accuracy (54.18%), spatial-temporal (47.08%), and spatial (43.58%). For happy expression classification conditions, spatial-temporal-ordinal accuracy had the highest accuracy (55.35%%), followed by spatial-temporal (49.32%), and spatial (47.96%). For sad expression classification conditions, spatial-temporal-ordinal had the highest accuracy (55.55%), spatial-temporal (47.08%), and spatial (46.98%). When looking at the spatial-temporal image classification conditions, happy (49.32%) had the highest accuracy, followed by angry (47.76%), followed by fear and sad (47.08%, respectively). When examining the spatial image classification conditions, happy (47.96%) had the highest accuracy, followed by sad (46.98%), and angry and fear (43.58%, respectively).

6.2.9. Decoding Results – Cross Decoding (High-A vs. Low-A)

To examine the influence of alexithymia severity on decoding accuracy, the classifiers were retrained and tested. For these new classification sets, all previous conditions were held the same, but the training and testing data fed into the classifier were modified. Classifiers were instead trained on ASD data with the inputs being High-A and Low-A individuals and tested on NT data for High-A and Low-A individuals. This form of cross classification allowed us to investigate any underlying featural similarities more directly for eye movement behaviors during facial expression processing when individuals were matched on alexithymia severity. If our prediction that eye movement patterns of ASD High-A individuals are similar to NT High-A individuals, and that eye movement patterns of ASD Low-A individuals are similar to NT Low-A individuals, then the classifier should perform significantly above chance level during cross classification.

Table 2 shows the decoding accuracy for all the classification conditions for the ASD High-A /Low-A training data and NT High-A/ Low-A testing data. Results indicate that the classification conditions for angry spatial images (54.67%), fear spatial images (55.64%), happy spatial-temporal-ordinal (54.67%), spatial-temporal (54.09%) and spatial images (54.48%) performed significantly above chance level ($p < .001$, respectively). No other classification conditions reached significantly above-chance level decoding accuracy. Regarding ESE on overall decoding accuracy, eye movement patterns related to the processing of fear facial expressions at the spatial level yielded the highest accuracy compared to all other expressions and scanpath information. These results were observed when the classifier was fed with more relevant scanpath information being plotted onto the images (i.e., the Spatial-Temporal-Ordinal images). Results indicated that the previously observed increase in decoding information when more scanpath information was plotted onto the images and fed into the classifier was only observed in the happy condition, where happy spatial-temporal-ordinal had the highest accuracy, however spatial was the next highest, followed by spatial-temporal having the lowest. For all other expression conditions, it was observed that the spatial scanpath images yielded the highest decoding accuracy within emotional category. For the angry classification condition, angry spatial images (54.67%) yielded the highest accuracy followed by angry spatial-temporal images (49.22%), and angry spatial-temporal-ordinal images (45.91%). This same pattern of spatial images resulting in the highest accuracy was also observed for fear, where fear spatial images (55.64%) yielded the highest accuracy, followed by fear spatial-temporal-ordinal images (49.81%), and fear spatial-temporal images (48.25%). Sad classification conditions also followed this pattern, where sad spatial images (49.61%) yielded the highest accuracy, followed by sad spatial-temporal images (49.03%), and sad spatial-temporal-ordinal images (48.05%) yielding the lowest accuracy. This decoding behavior was again observed when all emotional expression information (angry, fear, happy, and sad) was fed into the classifier, where all expression decoding for all expressions spatial images (53.44%) yielded the highest accuracy, followed by all expressions spatial-temporal images (51.46%), and all expressions spatial-temporal-ordinal images (50.58%).

Figure 19 shows the confusion matrices for classification accuracy and confusion percentages for all angry classification conditions during our cross decoding of training the CNN on High-A and Low-A scanpath images of ASD individual, and testing on High-A and Low-A scanpath images of NT individuals. Figure 19(a) reveals that when spatial-temporal-ordinal scanpath images for angry were fed into the classifier, the classifier performed more accurately for Low-A individuals (54.86%) than for High-A (36.96%). Figure 19(b) reveals that when spatial-temporal scanpath images for angry were fed into the classifier, the classifier performed more accurately for Low-A individuals (53.70%) than for High-A (44.75%). Figure 19(c) reveals that when spatial scanpath images for angry were fed into the classifier, the classifier performed more accurately for High-A individuals (58.37%) than for Low-A (50.97%).

Figure 20 shows the confusion matrices for classification accuracy and confusion percentages for all fear classification conditions during our cross decoding of training the CNN on High-A and Low-A scanpath images of ASD individual, and testing on High-A and Low-A scanpath images of NT individuals. Figure 20(a) reveals that when spatial-temporal-ordinal scanpath images for fear were fed into the classifier, the classifier performed more accurately for Low-A individuals (56.03%) than for High-A (43.58%). Figure 20(b) reveals that when spatial-temporal scanpath images for fear were fed into the classifier, the classifier performed more accurately for High-A individuals (54.09%) than for Low-A (42.41%). Figure 20(c) reveals that when spatial scanpath images for fear were fed into the classifier, the classifier performed more accurately for Low-A individuals (56.03%) than for Low-A (55.25%).

Figure 21 shows the confusion matrices for classification accuracy and confusion percentages for all happy classification conditions during our cross decoding of training the CNN on High-A and Low-A scanpath images of ASD individual, and testing on High-A and Low-A scanpath images of NT individuals. Figure 21(a) reveals that when spatial-temporal-ordinal scanpath images for happy were fed into the classifier, the classifier performed more accurately for High-A individuals (73.15%) than for High-A (36.19%). Figure 21(b) reveals that when spatial-temporal scanpath images for happy were fed into the classifier, the classifier performed more accurately for High-A individuals (42.41%) than for High-A (44.75%). Figure 21(c) reveals that when spatial scanpath images for happy were fed into the classifier, the classifier performed more accurately for High-A individuals (58.75%) than for Low-A (50.19%).

Figure 22 shows the confusion matrices for classification accuracy and confusion percentages for all sad classification conditions during our cross decoding of training the CNN on High-A and Low-A scanpath images of ASD individual, and testing on High-A and Low-A scanpath images of NT individuals. Figure 22(a) reveals that when spatial-temporal-ordinal scanpath images for sad were fed into the classifier, the classifier performed more accurately for High-A individuals (67.32%) than for Low-A (28.79%). Figure 22(b) reveals that when spatial-temporal scanpath images for sad were fed into the classifier, the classifier performed more accurately for Low-A individuals (50.19%) than for High-A (47.86%). Figure 22(c) reveals that when spatial scanpath images for sad were fed into the classifier, the classifier performed more accurately for High-A individuals (64.59%) than for Low-A (34.63%).

Figure 23 shows the confusion matrices for classification accuracy and confusion percentages for the all-emotions classification conditions (i.e., when all scanpath images for angry, fear, happy, and sad were fed into the classifier) during our cross decoding of training the CNN on High-A and Low-A scanpath images of ASD individual, and testing on High-A and Low-A scanpath images of NT individuals. Figure 23(a) reveals that when spatial-temporal-ordinal scanpath images for all emotional expressions were fed into the classifier, the classifier performed more accurately for Low-A individuals (55.38%) than for High-A (45.78%). Figure 23(b) reveals that when spatial-temporal scanpath images for all emotional expressions were fed into the classifier, the classifier performed more accurately for High-A individuals (55.77%) than for Low-A (47.14%). Figure 23(c) reveals that when spatial scanpath images for all emotional expressions were fed into the classifier, the classifier performed more accurately for High-A individuals (57.90%) than for Low-A (48.98%).

6.2.10. Results Summary

In summation, our results suggest potential differences in the features of eye movement patterns for facial expression processing between ASD and NT individuals when accounting for alexithymia. We also found potential differences in eye movement patterns related to facial expression processing between High-A and Low-A individuals. However, these differences were mainly examined when using machine learning models and not statistical analyses such as T-tests. Additionally, contrary to our predictions, High-A individuals performed better at facial expression recognition when lexical information was present to categorize facial expression stimuli presentations, and these same behavioral results were also observed for Low-A individuals. Finally, we also found that ASD individuals performed significantly worse for FER at lower intensities of expressions during the LLE task. Together these results may suggest that ASD and NT individuals have similar levels of FER performance even when accounting for alexithymia, but the observed differential eye movement features may suggest compensatory mechanisms of eye movement patterns related to facial expression processing.

CHAPTER 7. DISCUSSION

The goals of the current experiment were to examine if the deficits in FER performance often reported in ASD may be better explained by comorbid alexithymia, and to investigate if alexithymia FER performance may be influenced by abnormal recognition processes involving verbal labeling of facial expressions rather than by atypical perceptual processing of facial expression. We measured individuals eye movement patterns to investigate the following hypotheses. First, we hypothesized that if the deficits in FER performance in ASD stems from alexithymic traits, there could be differential eye movement patterns between ASD individuals with higher alexithymia-severity scores compared to ASD individuals with lower alexithymia-severity scores. Previous research has found that High-A and Low-A individuals have differing eye movement patterns in relation to face processing, such that High-A individuals are more likely to attend to the mouth region of a face than Low-A individuals (Fujiwara et al., 2018). If so, it is predicted that eye movement patterns of high-alexithymic ASD individuals would be similar to high-alexithymic NT individuals and that low-alexithymic ASD individuals would have eye movement patterns comparable to low-alexithymic NT individuals. Our results provide support for this first hypothesis, such that we found differences in eye movement patterns between ASD and NT individuals, as well as no significant differences in eye movement patterns between ASD High-A and NT High-A individuals. However, it should be stated that although we found significant differences between ASD High-A and NT High-A individuals, these results may better describe eye movement behaviors related to unconscious gaze processes, rather than eye movement behaviors related to face processing because the fixation and dwell times were small for these groups. Second, we also hypothesized that if alexithymic traits in individuals were based on recognition and labeling processes rather than perceptual processes of facial expressions, individuals with higher levels of alexithymia would perform worse than individuals with lower levels of alexithymia at FER when tasked with using lexical labels to categorize expressions compared to the facial expression matching task. However, our findings do not support this second hypothesis. Instead, we found that all individuals performed best at FER when lexical labels of expressions were present rather than schematic representations of faces, suggesting that lexical labels of expression may be the preferred response option during FER tasks.

7.1. Eye Movement Related Differences

Our results showed a significant difference in eye movement patterns for facial expression processing in that High-A individuals made more fixations and longer durations of fixations to the nose than Low-A. These results may suggest an alexithymic-specific effect on how alexithymia influences the looking behaviors for internal facial features, such that High-A individuals may engage with internal facial features differently than their Low-A counterparts. These findings also support previous work from Fujiwara et al. (2018), who found that High-A individuals were more likely to attend to certain internal facial features than Low-A individuals. However, it should be noted that Fujiwara et al. (2018) found that High-A individuals were more likely to attend to the mouth than Low-A individuals, and we found that High-A individuals were only more likely to attend to the nose than Low-A individuals. Additionally, our results indicate that

ASD Low-A made more fixations to screen background and for longer than NT Low-A. These results would be consistent with prior work such as Kang et al. (2020), which suggests that ASD individuals spend more time looking away from the subject of the stimulus presentation. However, our fixation durations regarding these results were below our established 60 ms fixation duration threshold, which may suggest that our results were instead due to unconscious gaze processes rather than facial expression processing behaviors due to both our fixation counts as well as the dwell times being so low. Additionally, other than potential differences in unconscious gaze processes, we found ASD and NT had similar eye movement patterns overall. These results may suggest that ASD and NT individuals may engage with faces in a similar manner during facial expression processing, such that they do not vary in how often or for how long they look at various AOI's, such as the presentation screen, image area, face, eyes, left eye, right eye, mouth, and nose.

Although we found minimal group differences when running statistical analyses for fixation information (AOI fixation count and dwell times), we found that we could significantly decode above chance level for ASD and NT individuals while controlling for alexithymia severity, as well as during cross-decoding methods where we trained the classifier on High-A and Low-A scanpath images from ASD individuals and tested on High-A and Low-A scanpath images from NT individuals. We fed our classifier with various scanpath image plots differing on which information was plotted onto the image (spatial, spatial-temporal, or spatial-temporal-ordinal). A shortcoming of neural networks is that gaining insight into what exact features lead to accurate decoding becomes impossible. However, our image manipulations allowed us to further examine at what level of eye movement behaviors do group differences potentially emerge. Our results showed that we could decode significantly above chance level for all scanpath image information conditions as well as all emotional expression conditions when attempting to decode for ASD High-A, ASD Low-A, NT High-A, and NT Low-A individuals. Additionally, these results showed that as more scanpath image information was fed into the classifier, the greater decoding accuracy became. Together, these results may suggest that there are present featural differences of eye movement behaviors in relation to facial expression processing at the spatial, spatial-temporal, and spatial-temporal-ordinal level for these groups and that machine learning methods could be more sensitive to detecting differences in facial expression processing in relation to eye movements than conventional statistical methods. Additionally, our cross-decoding results showed that when we used scanpath images of spatial eye movement information, we could significantly decode above chance level for each of the emotional decoding conditions except for sad. Additionally, when the classifier was fed with spatial-temporal or spatial-temporal-ordinal scanpath images, the classifier was no longer performing significantly above chance level. This may mean that when temporal and ordinal information is introduced to the classifier, there may be too much noise or similarities between the groups to accurately classify which group an individual may belong to. These results suggest that there may be distinct eye movement behaviors for facial expression processing when comparing between High-A and Low-A individuals at the spatial level.

7.2. FER Performance

Contrary to our prediction that High-A individuals would perform better in the pictorial matching of expressions than in the lexical labeling of expressions, we found that High-A individuals performed significantly worse in the FEM task than in the LLE task. However, we also found these same behavioral patterns in Low-A individuals. These results may suggest that this observed performance when no lexical information was present may instead be due to task difficulty. For instance, it may be that individuals are more regularly exposed to emotional labels of expressions than schematic presentations of expressions; thus, labels may be more ecologically valid than schematic expressions when being asked to categorize a facial expression. Additionally, although schematic faces of expressions display the main features of a face (e.g., Bi et al., 2022), we did not match our schematic expressions to the intensities of the realistic pictorial facial expression stimuli. Our schematic expressions only portrayed one intensity level (full intensity), while our presentations of facial expressions were at 40%, 60%, 80%, and 100% intensities. A key component of micro-expressions is that they be low-intensity expressions (e.g., Lu et al., 2022) Therefore, it could be that this mismatch for intensities amongst realistic presentations and schematic responses made it difficult for individuals to match these potential micro-expressions or lower-intensity realistic expressions to their full-intensity schematic face counterparts. However, worse FER performance during the FEM task may also suggest that lexical labeling of expressions may be a preferred method for behavioral response inputs during facial expression recognition tasks.

Comparisons between ASD and NT showed that ASD individuals only performed worse at FER at lower intensities. This lower FER performance at lower intensities for ASD individuals would support earlier findings of Griffiths et al. (2019), who also showed worse performance at lower intensities for ASD individuals. Additionally, these differences in behavioral performance suggest that ASD individuals may be less accurate at categorizing micro-expressions. Thus, ASD individuals may have greater difficulty in discerning these brief, subtle, and more ecologically common expressions (e.g., Dong et al., 2012), potentially making it harder for them to navigate more common aspects of social interactions. These minimal group differences between ASD and NT individuals regarding FER performance may be due to our ASD sample being potentially more high functioning, as we only showed a weak negative correlation between ASD and IQ, and research indicates that FER differences between NT and high-functioning ASD individuals may be less severe than previously thought (e.g., Kessels et al., 2010). However, it should also be noted that Leung et al. (2018) argue that inconsistencies in face processing ASD research may be due to numerous differences found between studies surrounding their inclusion and exclusion criteria or the paradigms used during experiments. It is possible that the nature of our two tasks, or our exclusion and inclusion criteria, may also be why we did not find any further group differences between ASD and NT individuals. Previous work has indicated that FER abilities for ASD individuals may be better for NT individuals (e.g., Evers et al., 2015), yet we only found group differences at lower intensities of expressions.

7.3. Limitations and Future Directions

It should be noted that the present study had a particularly low sample size of ASD individuals. To combat this limitation for our decoding analyses methods, such as using a BNCDF, were implemented (e.g., Combrisson & Jerbi, 2015), but no such corrections were applied for statistical analyses. Therefore, our study was likely underpowered, making it difficult to accurately interpret behavioral findings centered around ASD individuals. Additionally, our study tasked individuals with completing the

AQ-10, TAS-20, and WAIS-IV Matrix Reasoning task prior to performing either of our FER tasks. It could be possible that this caused potential response biases from participants if they were able to discern the meaning behind these measures (e.g., Posserud et al., 2010), such that an individual may perform worse during the FER tasks if they felt as if they performed poorly on our general IQ measure.

The present study also did not include more emotional expressions that could be categorized as positively valenced (e.g., joy). Therefore, future research may want to examine any valence-specific influence on decoding accuracy. For example, would researchers again see an increase in decoding accuracy during cross-decoding methods for a positively valenced expression compared to negatively valenced expressions, such as in our results where we could decode significantly above chance level for scanpath image plots of spatial-temporal and spatial-temporal-ordinal plots related to happy facial expressions but not for the negatively valenced expressions of angry, fear, and sad. Additionally, it may be interesting to examine the temporal dynamics of eye movement data related to facial expression processing to investigate when groups (either High-A vs. Low-A, or ASD High-A vs. ASD Low-A vs. NT High-A vs. NT Low-A) can be accurately differentiated from one another. For example, using a Monte Carlo Simulation consistent with Bae and Luck (2018) to examine when decoding results that reached above chance level significantly varied from randomly generated distributions of data.

Future studies may also wish to include different populations of ASD individuals. The present study only included ASD individuals attending college who may be considered more high-functioning. Prior literature has suggested that disruptions and deficits in face processing can vary at different ages, such as Black et al. (2017)

concluding that eye-related differences in face processing become more apparent as ASD individuals get older, and that low-functioning ASD individuals may perform worse than their high-functioning ASD counterparts (e.g., Harms et al., 2010). It is possible that including ASD individuals from various populations may reveal further differences in FER performance for ASD individuals compared to NT individuals. Finally, although our results may suggest that FER performance may be increased when lexical labels are available to make behavioral responses for facial expression presentations, it may be interesting to examine the influence of a "hybrid" response option for expressions. For example, rather than showing lexical labels of expressions or schematic drawings of expressions, include a third task that displays schematic drawings with lexical labeling either above or below the drawing. It could be that adding both verbal and nonverbal information for behavioral response options may help alexithymic individuals overcome any potential deficits in understanding verbal and nonverbal representations of emotions.

7.4. Conclusion

By using eye movement decoding techniques coupled with presentations of facial expressions to participants, we were able to garner further insight into potential perceptual processing differences of facial expressions for individuals with autism spectrum disorder and neurotypical individuals and their alexithymia severity levels, such that the influence of alexithymia on eye movement patterns for facial expression processing may manifest themselves at the spatial level. This better understanding of how facial expressions are potentially differentially perceived and encoded could open the door for future studies to examine other influences of the perception and encoding of face-related information for various diagnostic groups.

Despite the limitations of our study, the current experiment demonstrated further support for there likely being differences in the eye movement patterns between individuals with autism spectrum disorder and neurotypical individuals (e.g., Black et al., 2018), as well as high-alexithymic and low-alexithymic individuals (e.g., Kang et al., 2020). Increased decoding performance in the high-alexithymic vs. low-alexithymic classification as seen when the Convolutional Neural Network was fed scanpath images containing spatial information for angry, fear, happy, and all emotional expressions (angry, fear, happy, and sad), but not sad, may suggest that there are in fact emotion-specific effects of eye movement patterns for alexithymia severity and that these differences manifest themselves at the spatial level. This demonstration of differences for eye movement decoding over statistical analyses in these specific classifying conditions may suggest that machine learning methods may be favorable at finding these eye movement-related differences due to their strengths of featural learning.

Additionally, consistent with past literature, we found that individuals with autism spectrum disorder performed significantly worse at facial expression recognition for lower intensities of expression. Thus, individuals with autism spectrum disorder may have greater difficulty in describing more ecologically valid micro-expressions. We also found that individuals with greater levels of alexithymia severity were less accurate in their categorizations of angry and sad faces but not happy and fearful faces. These results were found for both verbal and nonverbal categorizations of expressions and may suggest that the difficulty in describing emotional states observed in alexithymia may not be present for happy due to a potential positive classification advantage (e.g., Liu et al.,

2013) or for fear due to a potential fear-specific threat detection advantage (Hedger et al., 2015).

Overall, we suspect that these results may suggest that individuals, whether they belong to autism spectrum disorder, neurotypical, high-alexithymic, or low-alexithymic groups, can perform similarly to one another during facial expression tasks but may employ compensatory eye movement strategies when tasked with facial expression recognition. It is possible that these groups may utilize different looking behaviors during facial expression processing, but these differences may allow them to perform similarly to one another. Finally, these results may provide further support and additional information to be considered in the alexithymia hypothesis that the facial expression recognition deficits often observed in autism spectrum disorder may be better explained by frequently co-occurring alexithymia (e.g., Bird & Cook, 2013). Here, we suggest that if it is to be believed that alexithymia severity influences facial expression recognition, then the influence of alexithymia may manifest itself at the spatial level of eye movement behavior related to facial expression processing.

Figures

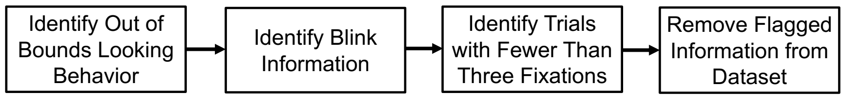


Figure 1. Data preprocessing flowchart.

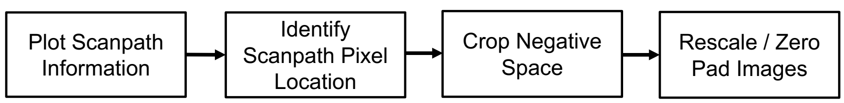


Figure 2. Image preparation flowchart.

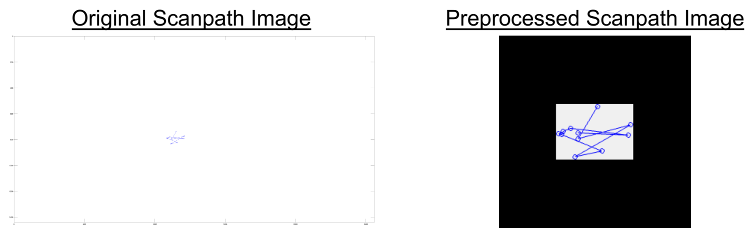


Figure 3. Representation of plotting scanpaths. On the left is the original scanpath, and on the right is the trimmed and resized scanpath to be used for later classification.

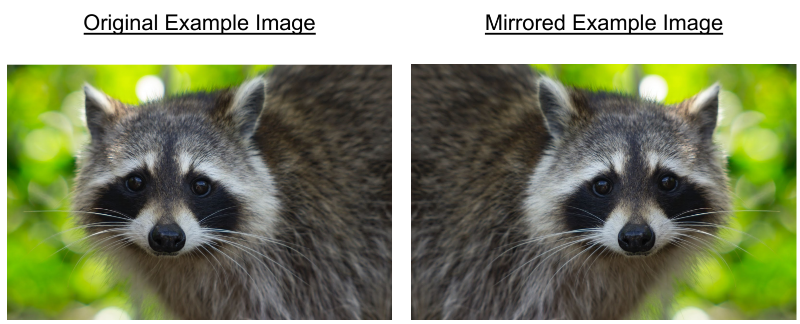
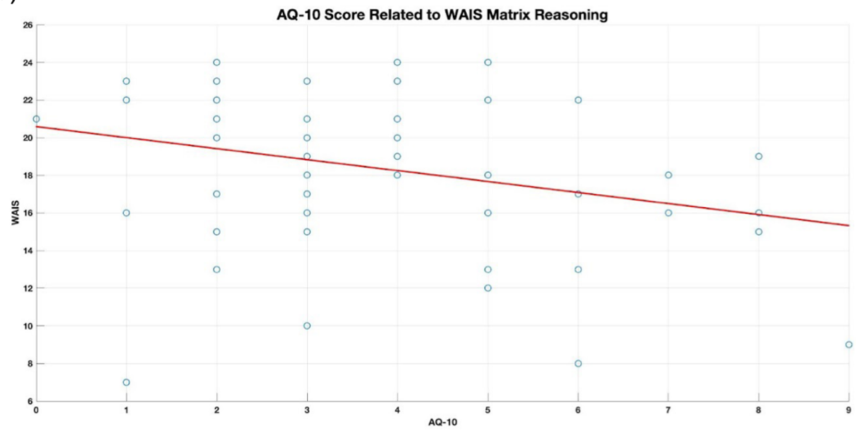
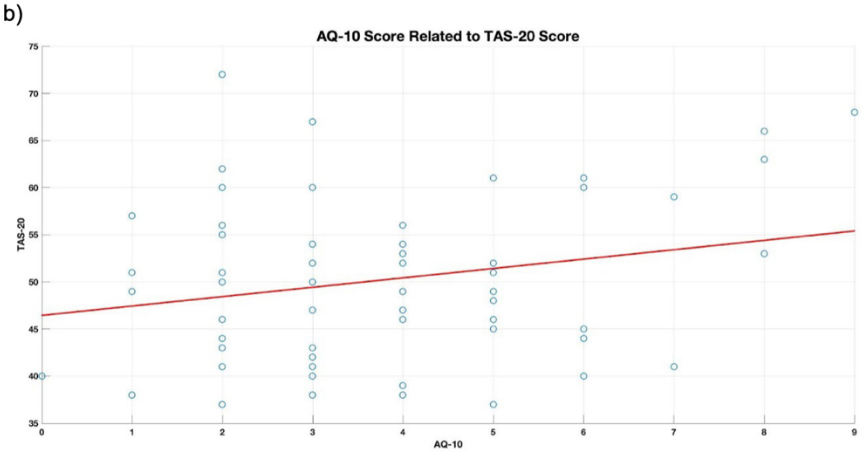


Figure 4. Representation of the data augmentation technique, mirroring. Here the original image of the raccoon is flipped along its horizontal axis to produce a new image that contains all the same underlying information.

a)



b)



c)

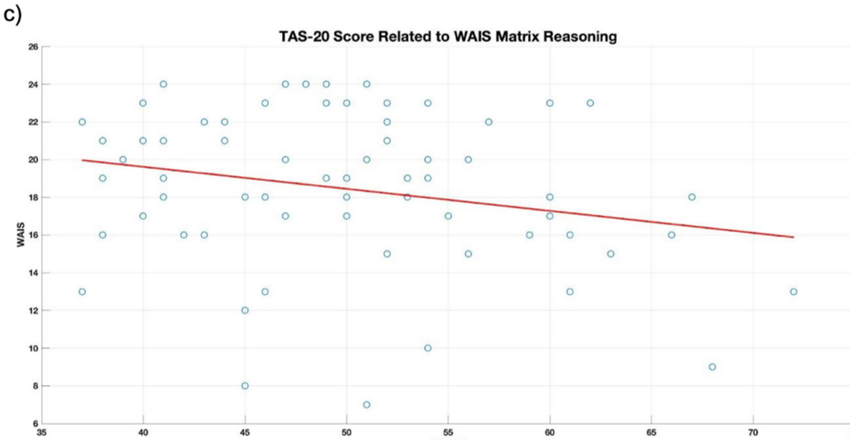


Figure 5. (a) Representation of the significant negative correlation (p = .03) between AQ-10 and WAIS-IV Matrix Reasoning scores. (b) Representation of the correlation

between AQ-10 and TAS-20 scores (p = .06). (c) Representation of the correlation between TAS-20 and WAIS-IV Matrix Reasoning scores (p = .06).

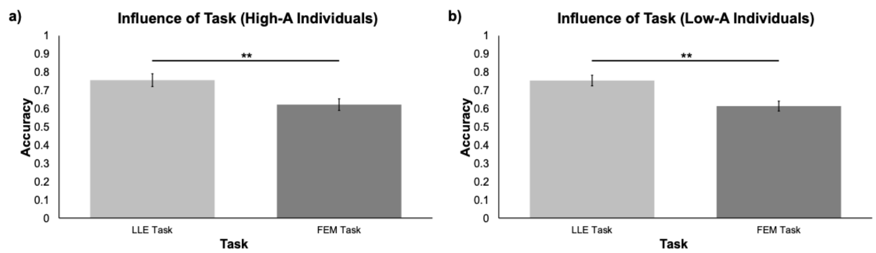


Figure 6. Bar graphs depicting differences in task performance. (a) Comparison of FER performance for the LLE and FEM tasks for High-A individuals. (b) Comparison of FER performance for the LLE and FEM tasks for Low-A individuals.

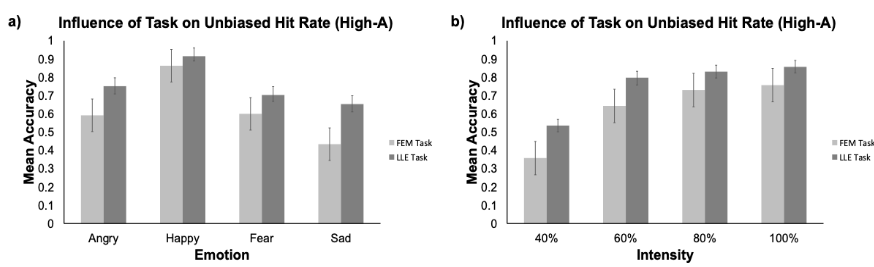


Figure 7. Bar graphs depicting differences in task performance for High-A individual based on emotional information. (a) Comparison of FER performance regarding emotional expression for the FEM and LLE tasks for High-A individuals. (b) Comparison

of FER performance regarding emotional expression intensity for the FEM and LLE tasks for High-A individuals.

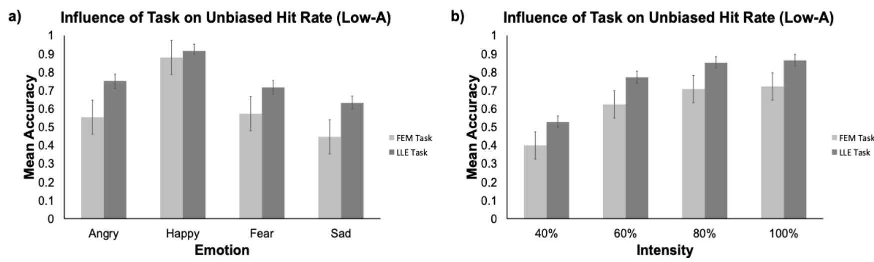


Figure 8. Bar graphs depicting differences in task performance for Low-A individual based on emotional information. (a) Comparison of FER performance regarding emotional expression for the FEM and LLE tasks for Low-A individuals. (b) Comparison of FER performance regarding emotional expression intensity for the FEM and LLE tasks for Low-A individuals.

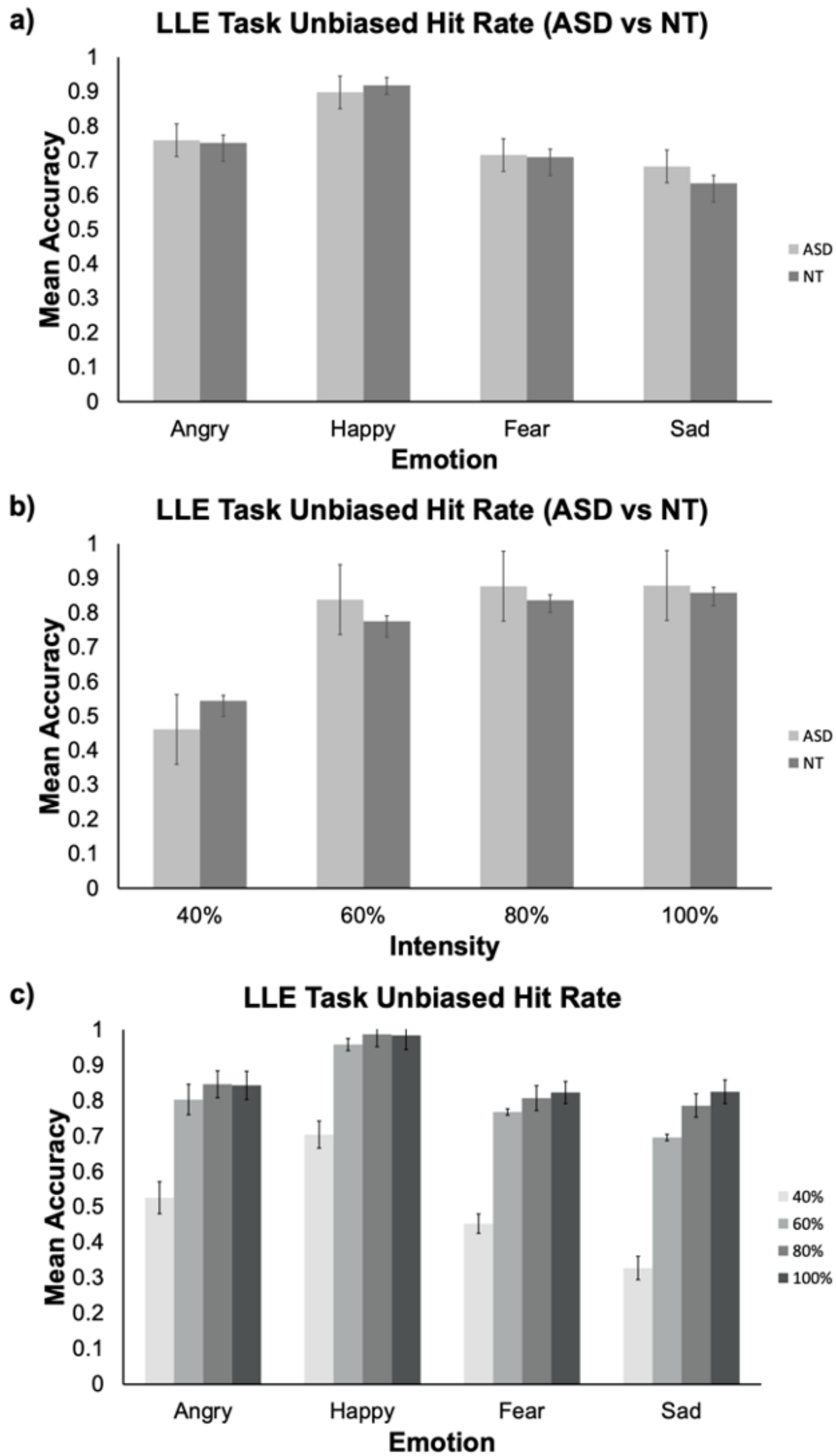


Figure 9. Bar graphs depicting performance of ASD and NT individuals during the Lexical Labeling of Expressions Task (a) Comparison of facial expression recognition

performance between ASD and NT individuals. (b) Comparison of facial expression recognition accuracy at each emotional intensity level between ASD and NT individuals. (c) Comparison of facial expression recognition accuracy for each expression at each emotional intensity level for both ASD and NT individuals.

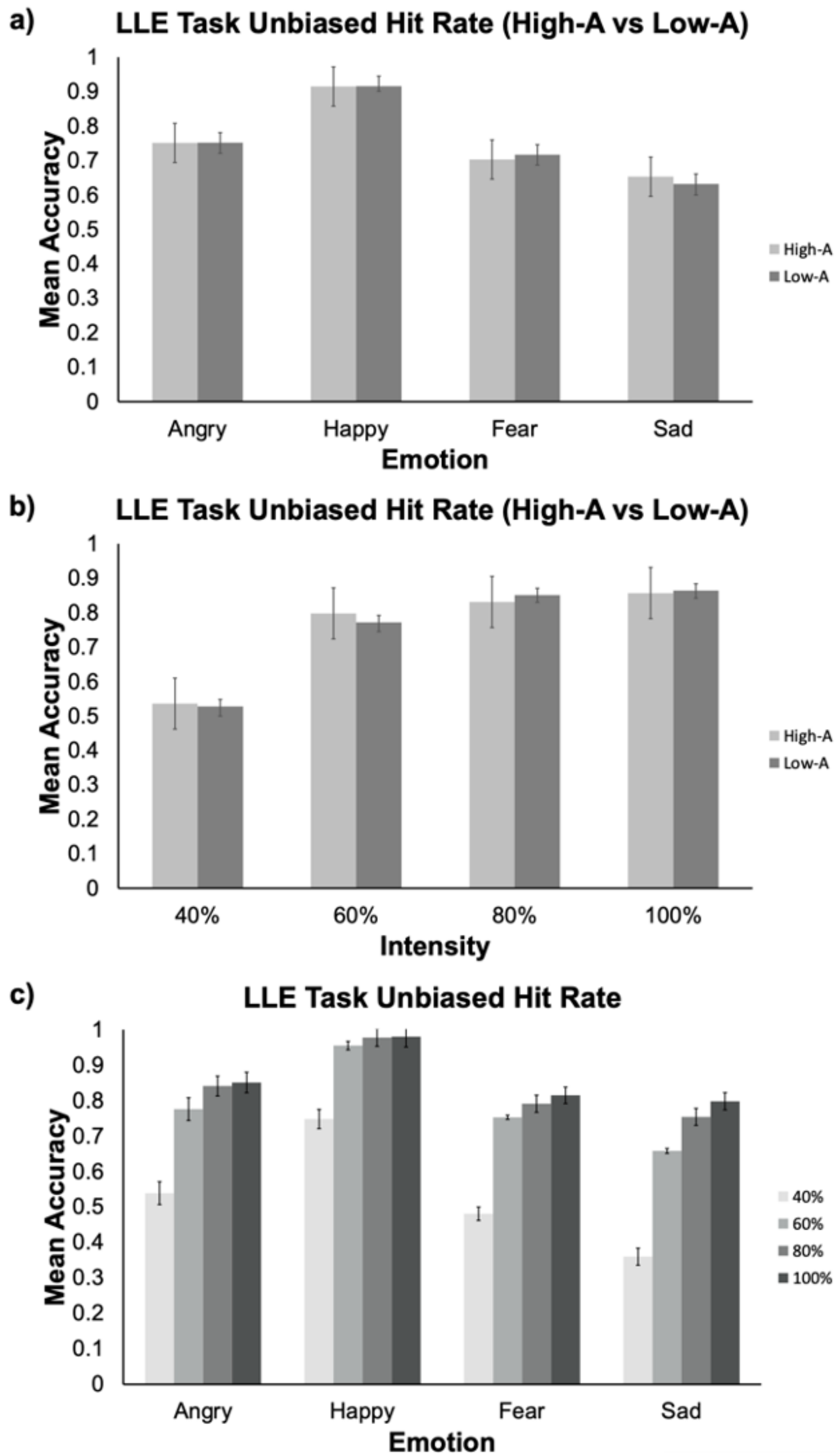


Figure 10. Bar graphs depicting performance of High-A and Low-A individuals during the Lexical Labeling of Expressions Task (a) Comparison of facial expression

recognition performance between High-A and Low-A individuals. (b) Comparison of facial expression recognition accuracy at each emotional intensity level between High-A and Low-A individuals. (c) Comparison of facial expression recognition accuracy for each expression at each emotional intensity level for both High-A and Low-A individuals.

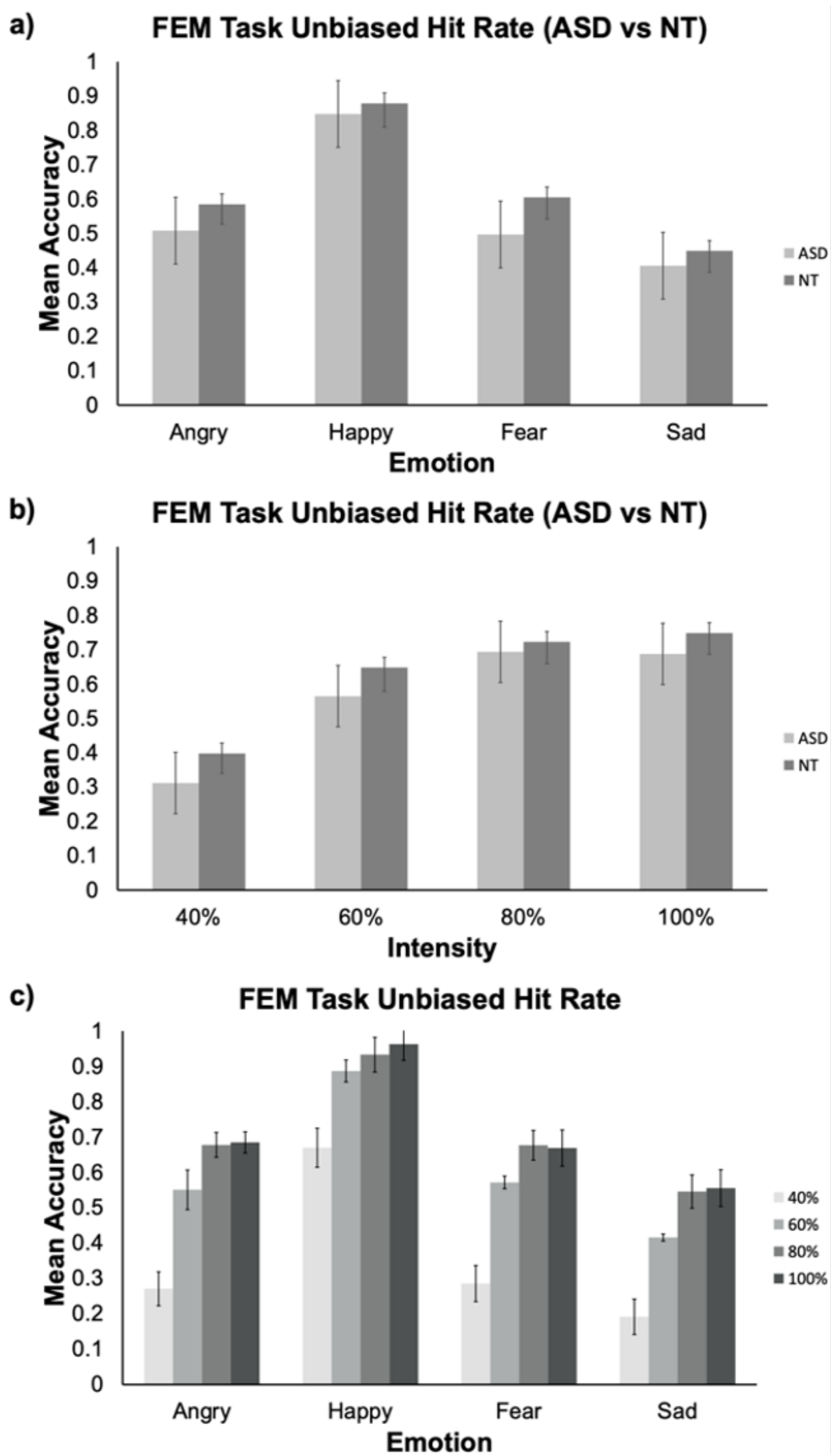


Figure 11. Bar graphs depicting performance of ASD and NT individuals during the Facial Expression Matching Task (a) Comparison of facial expression recognition

performance between ASD and NT individuals. (b) Comparison of facial expression recognition accuracy at each emotional intensity level between ASD and NT individuals. (c) Comparison of facial expression recognition accuracy for each expression at each emotional intensity level for both ASD and NT individuals.

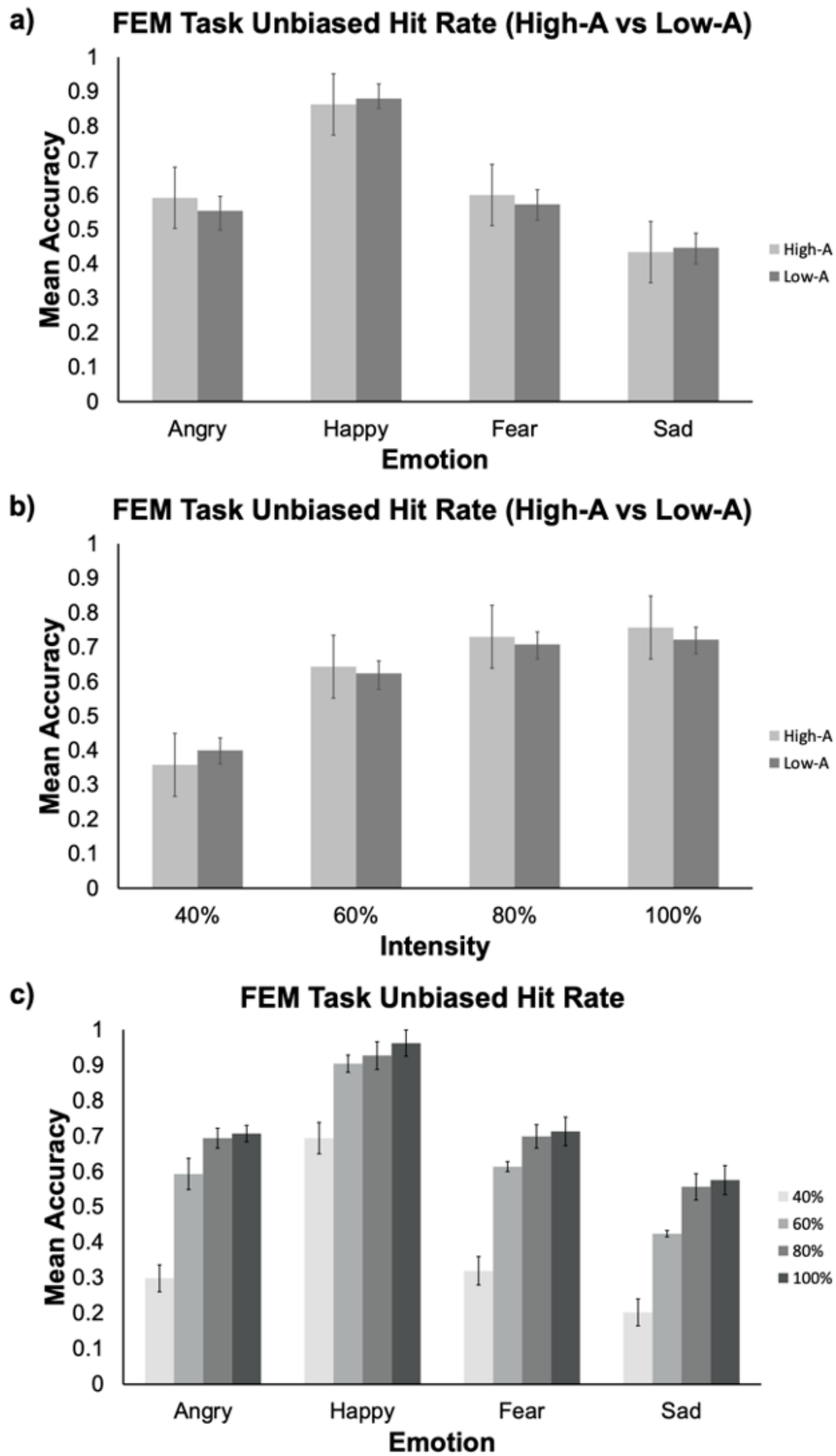


Figure 12. Bar graphs depicting performance of High-A and Low-A individuals during the Facial Expression Matching Task (a) Comparison of facial expression recognition

performance between High-A and Low-A individuals. (b) Comparison of facial expression recognition accuracy at each emotional intensity level between High-A and Low-A individuals. (c) Comparison of facial expression recognition accuracy for each expression at each emotional intensity level for both High and Low-A individuals.

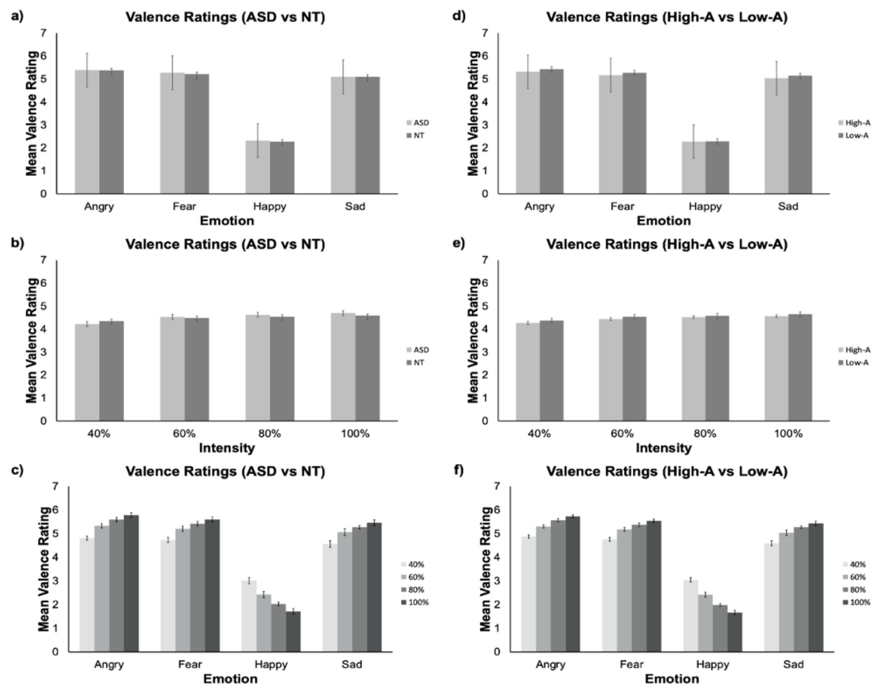


Figure 13. Bar graphs depicting valence ratings in response to facial expression stimuli and intensity of expression (a) Comparison of average valence rating between ASD and NT individuals for expressions. (b) Comparison of average valence rating between ASD and NT individuals for intensity. (c) Comparison of average valence rating between ASD and NT individuals for expression and intensity of expression. (d) Comparison of average valence rating between High-A and Low-A individuals for expressions. (e) Comparison of average valence rating between High-A and Low-A individuals for intensity. (f) Comparison of average valence rating between High-A and Low-A individuals for expression and intensity of expression.

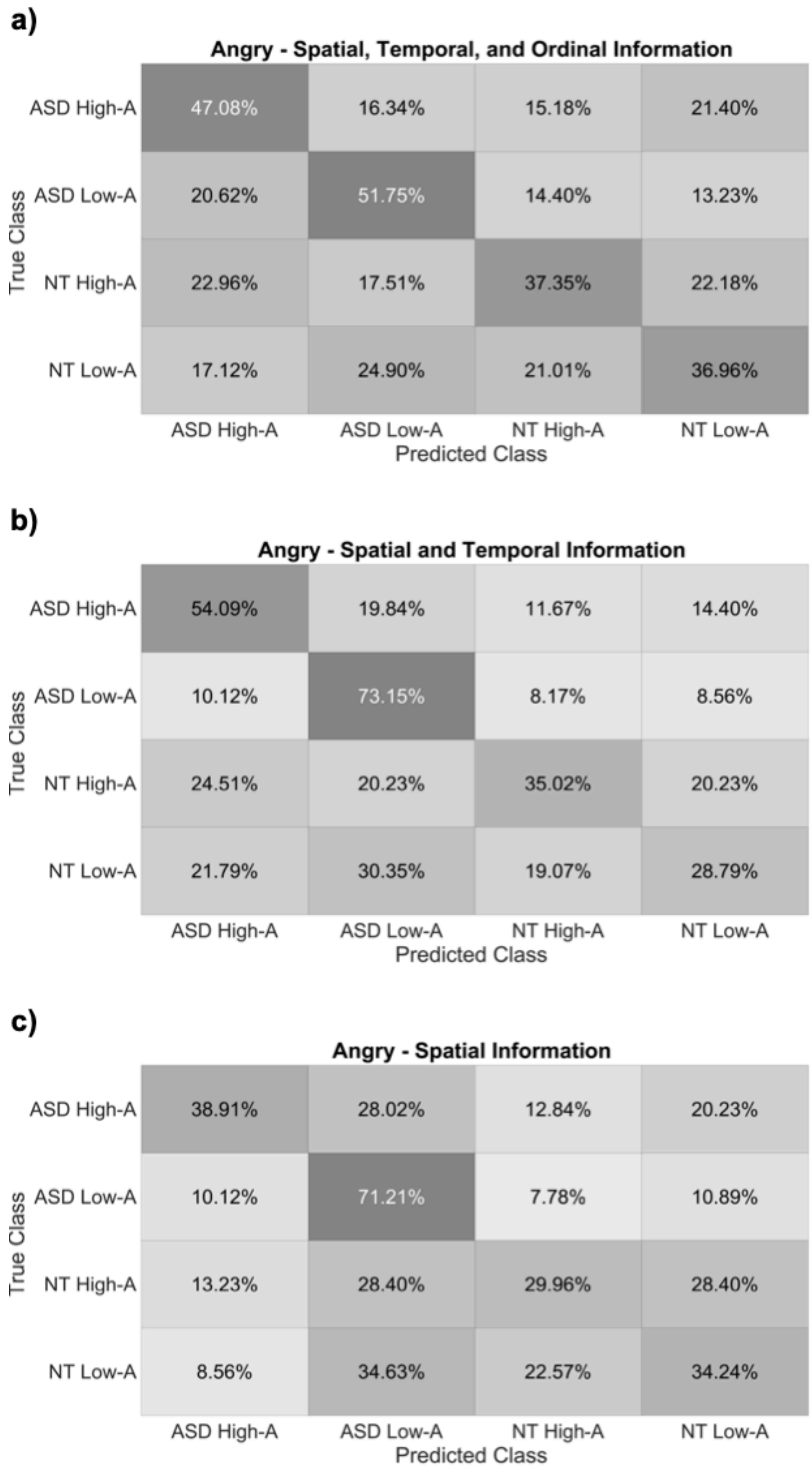


Figure 14. Confusion matrices showing classification accuracy and confusion percentages for all angry classification conditions. (a) Classification accuracy when using

Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

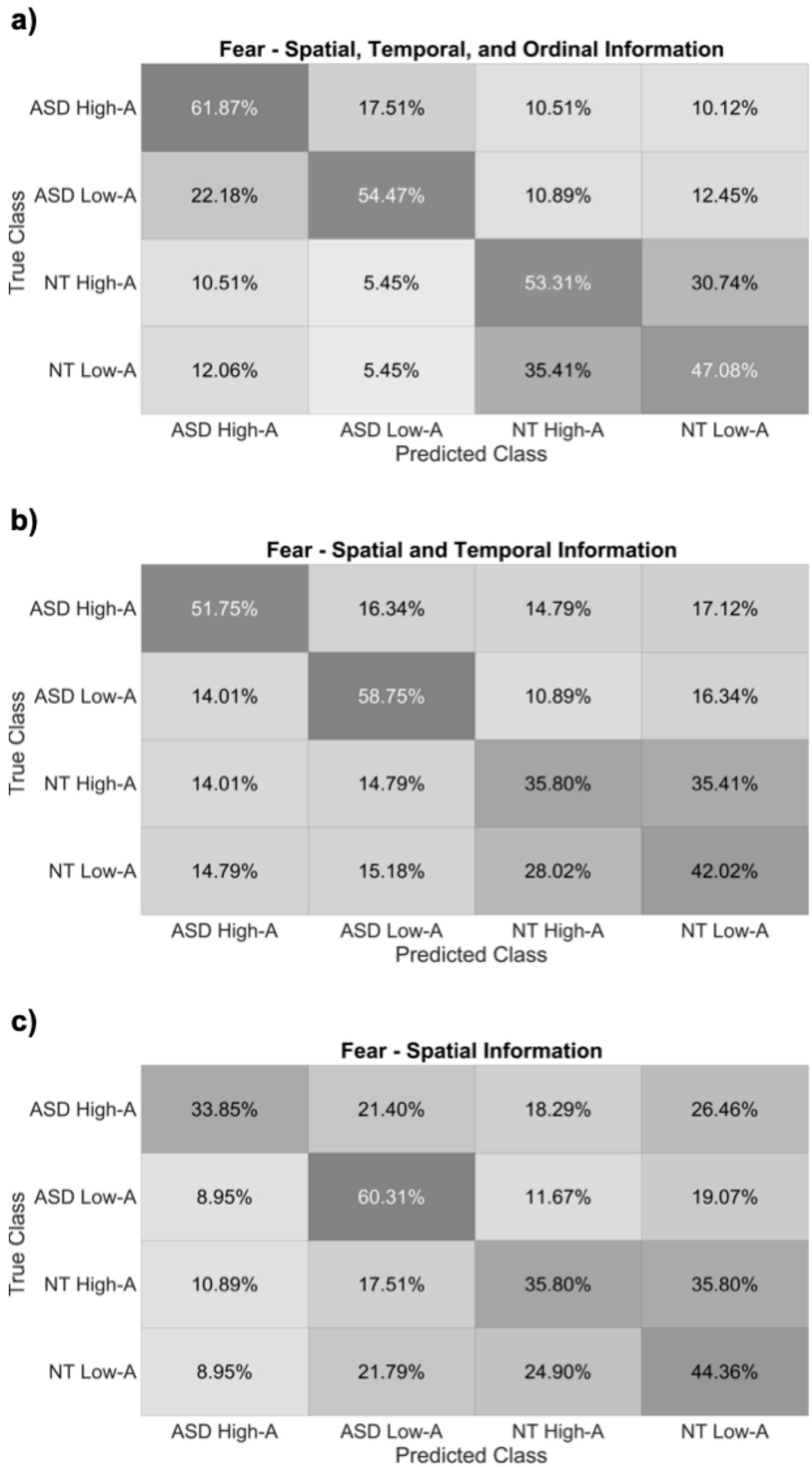


Figure 15. Confusion matrices showing classification accuracy and confusion percentages for all fear classification conditions. (a) Classification accuracy when using

Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

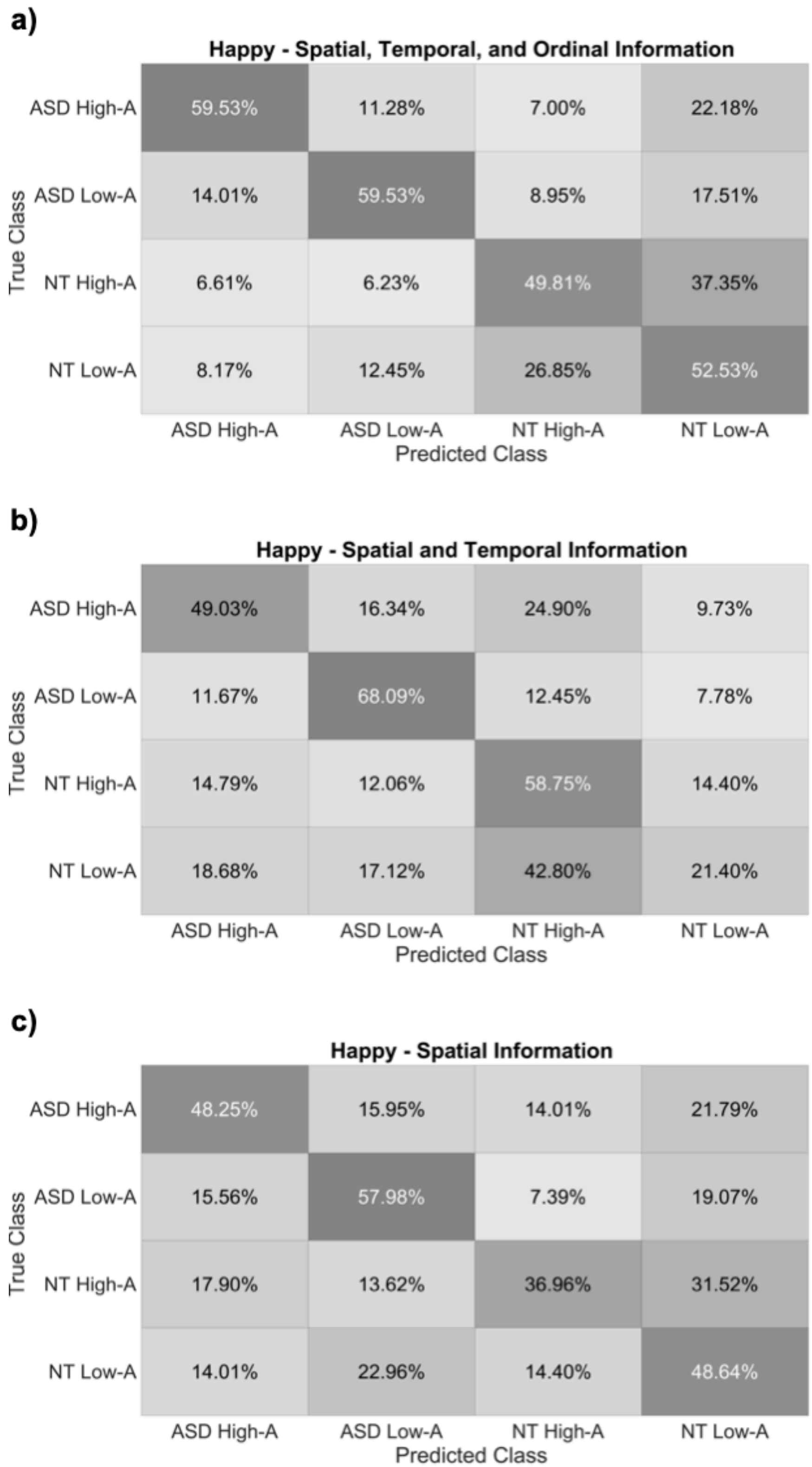


Figure 16. Confusion matrices showing classification accuracy and confusion percentages for all happy classification conditions. (a) Classification accuracy when

using Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

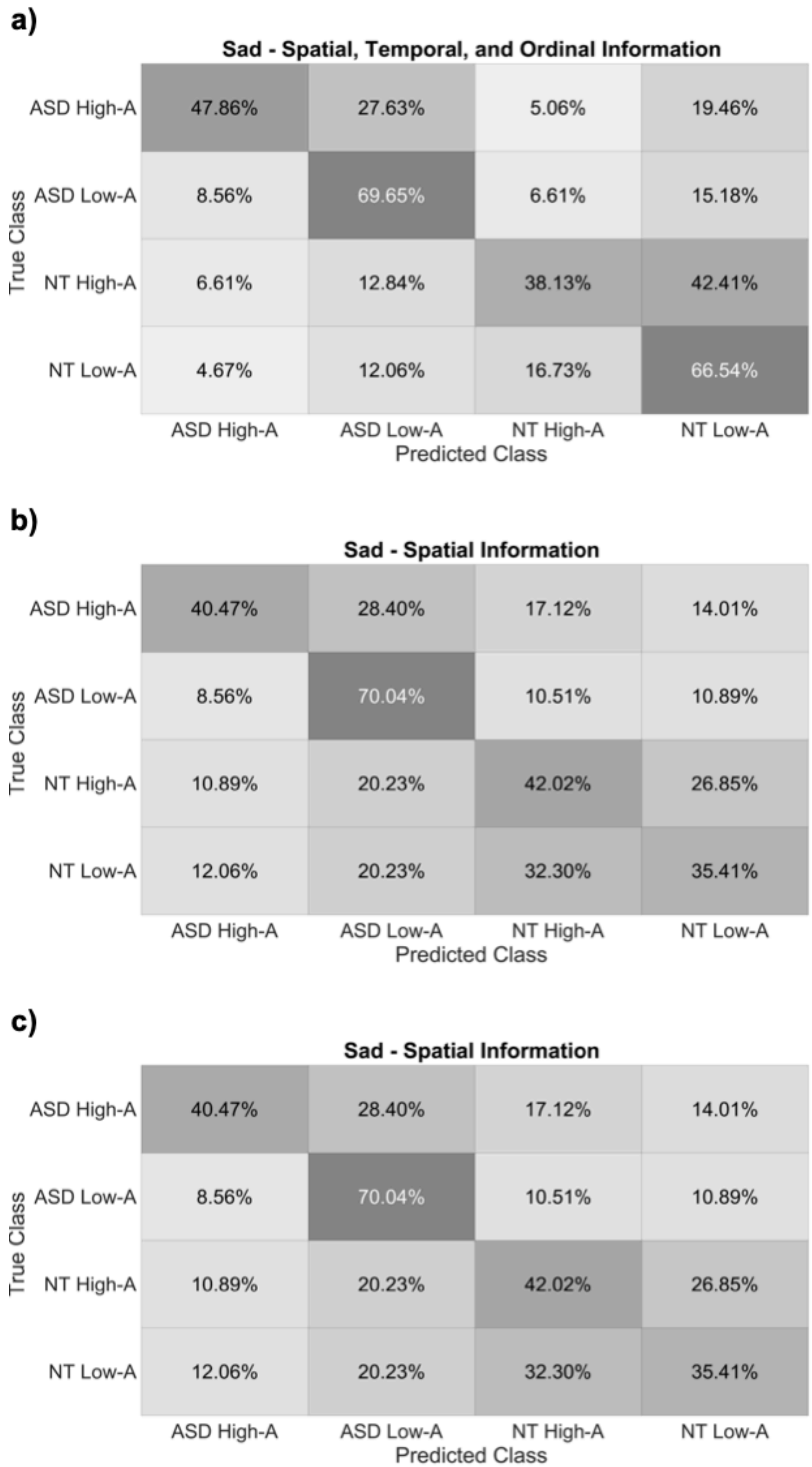


Figure 17. Confusion matrices showing classification accuracy and confusion percentages for all sad classification conditions. (a) Classification accuracy when using

Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

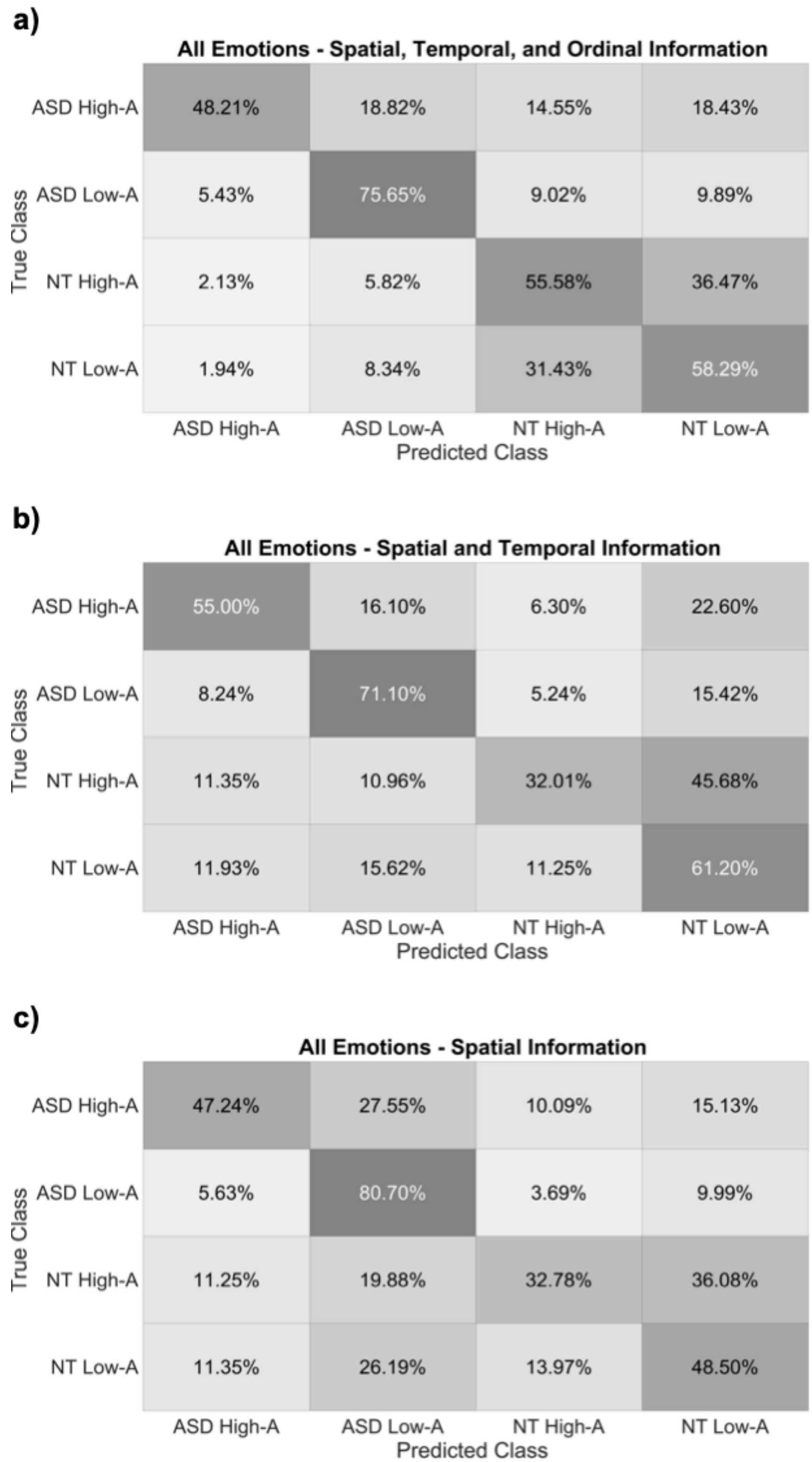


Figure 18. Confusion matrices showing classification accuracy and confusion percentages for all sad classification conditions. (a) Classification accuracy when using

Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

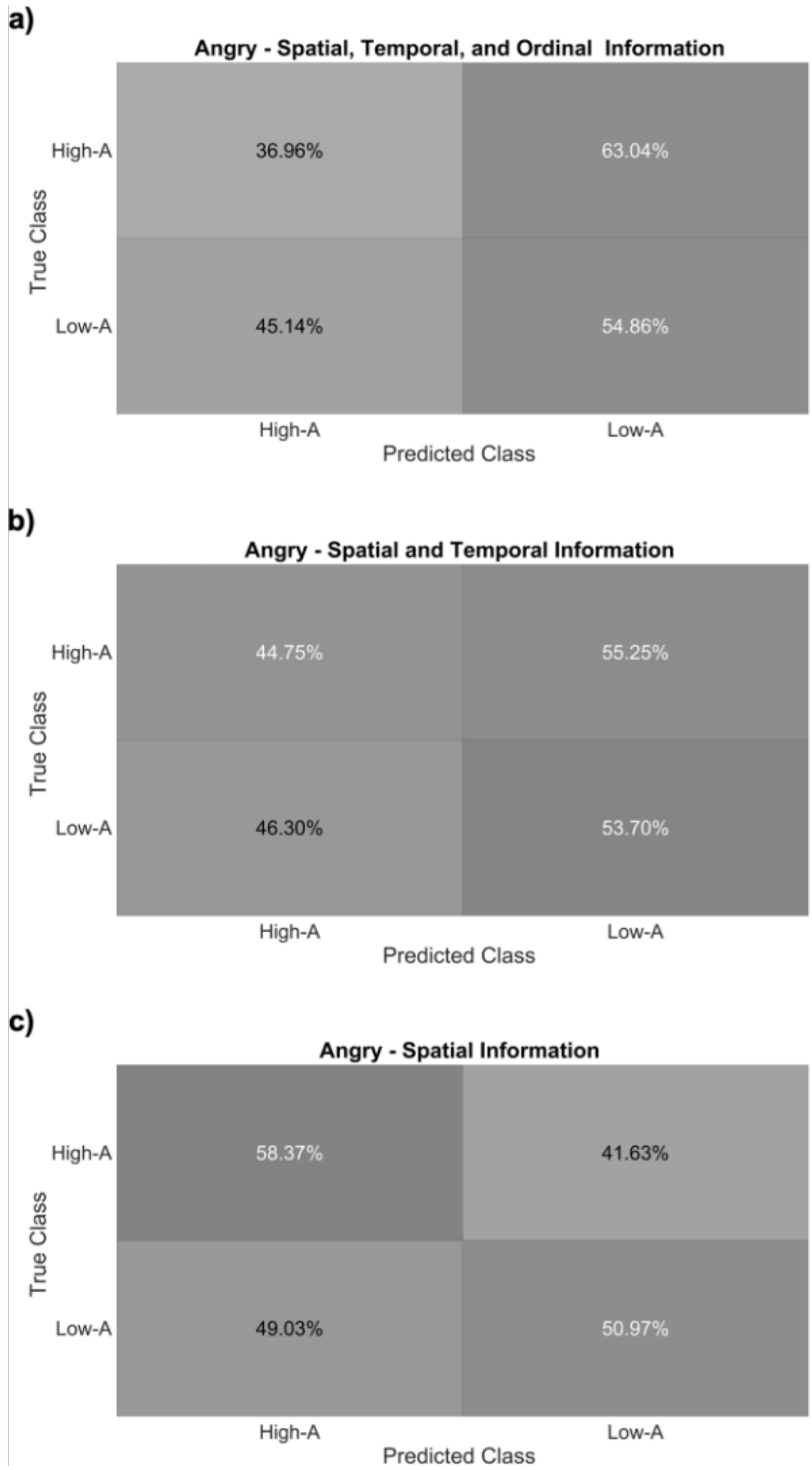


Figure 19. Confusion matrices showing classification accuracy and confusion percentages for all angry classification conditions when the classifier was trained on ASD

High-A/ Low-A data and tested on NT High-A/ Low-A data. (a) Classification accuracy when using Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

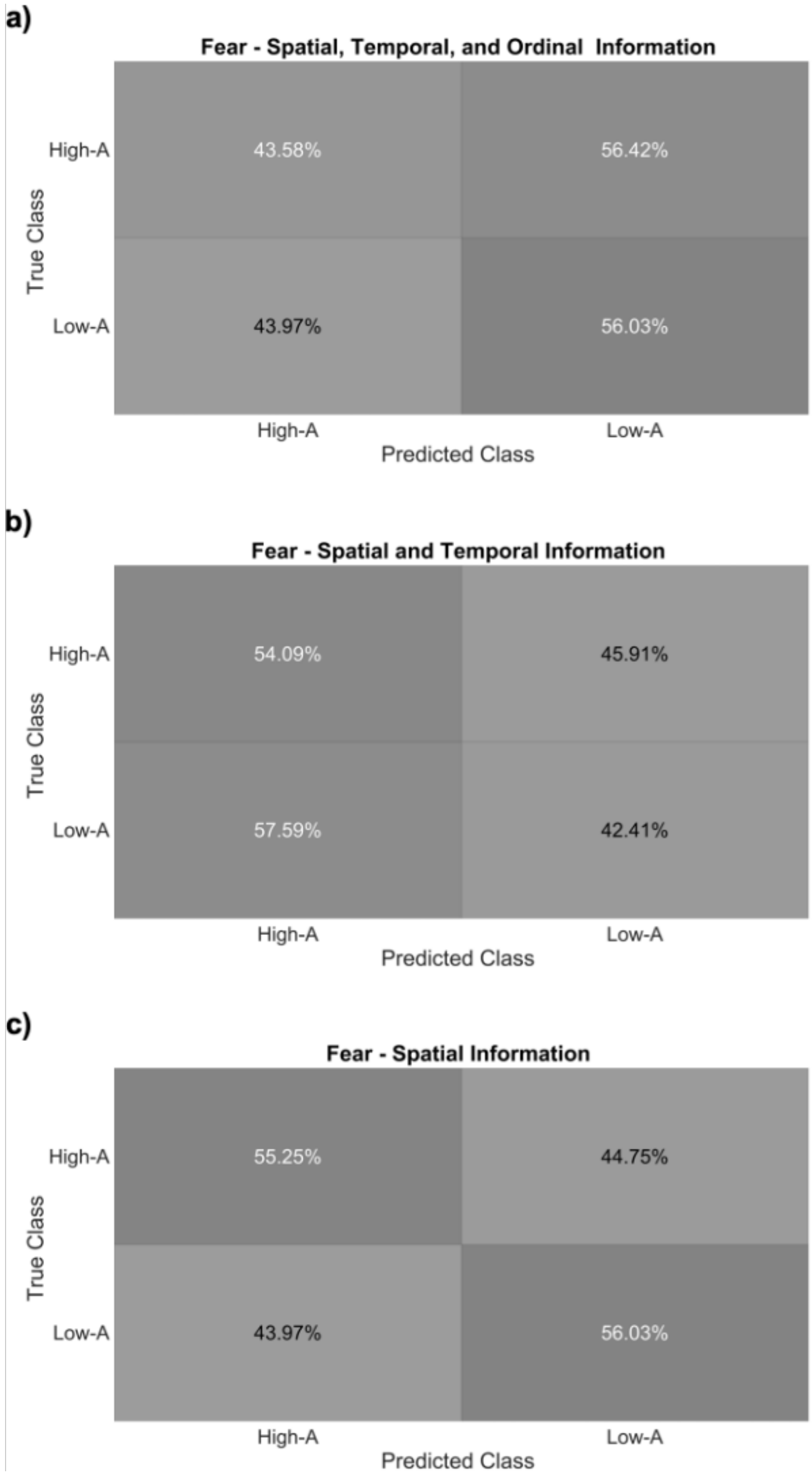


Figure 20. Confusion matrices showing classification accuracy and confusion percentages for all fear classification conditions when the classifier was trained on ASD

High-A/ Low-A data and tested on NT High-A/ Low-A data. (a) Classification accuracy when using Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

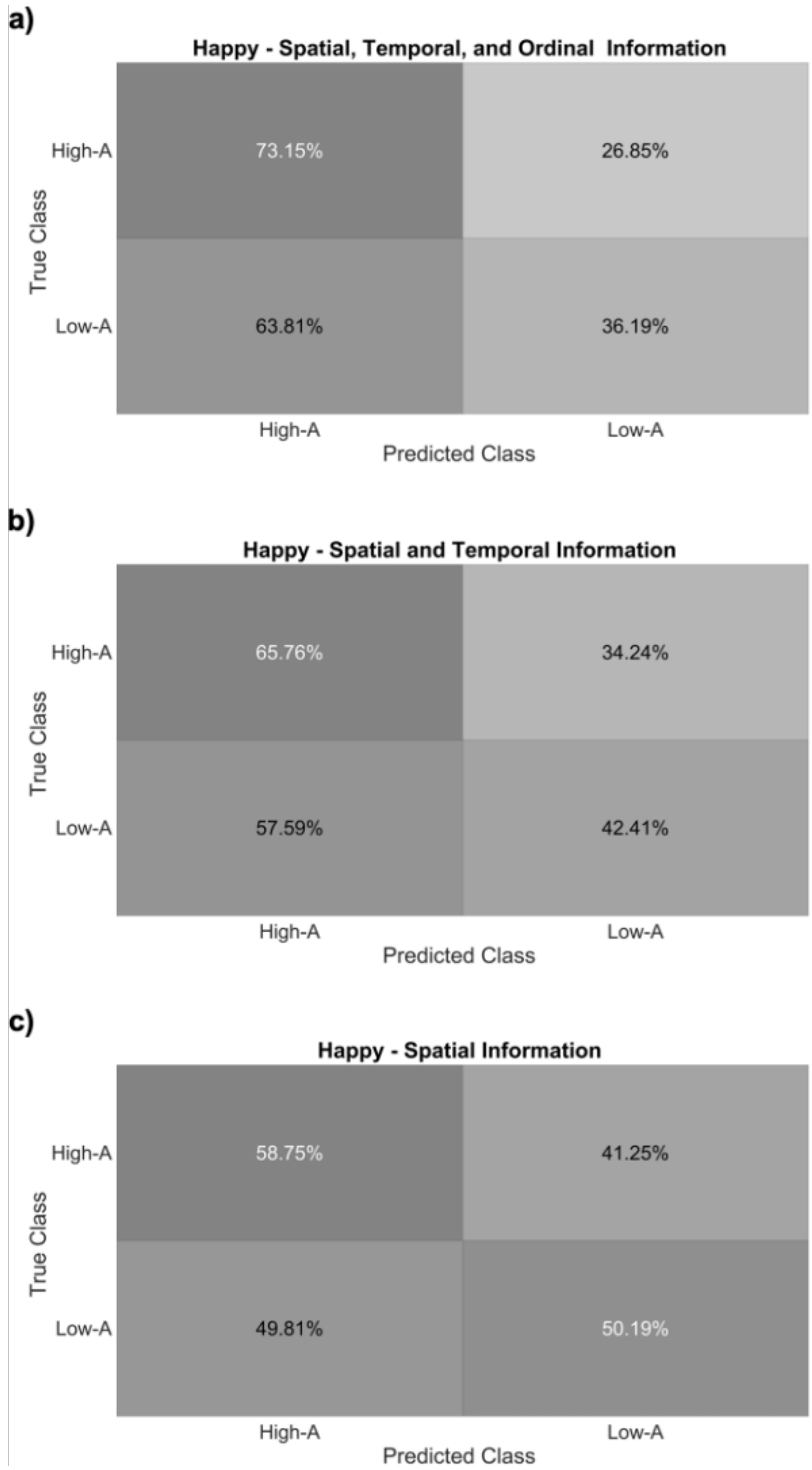


Figure 21. Confusion matrices showing classification accuracy and confusion percentages for all happy classification conditions when the classifier was trained on

ASD High-A/ Low-A data and tested on NT High-A/ Low-A data. (a) Classification accuracy when using Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

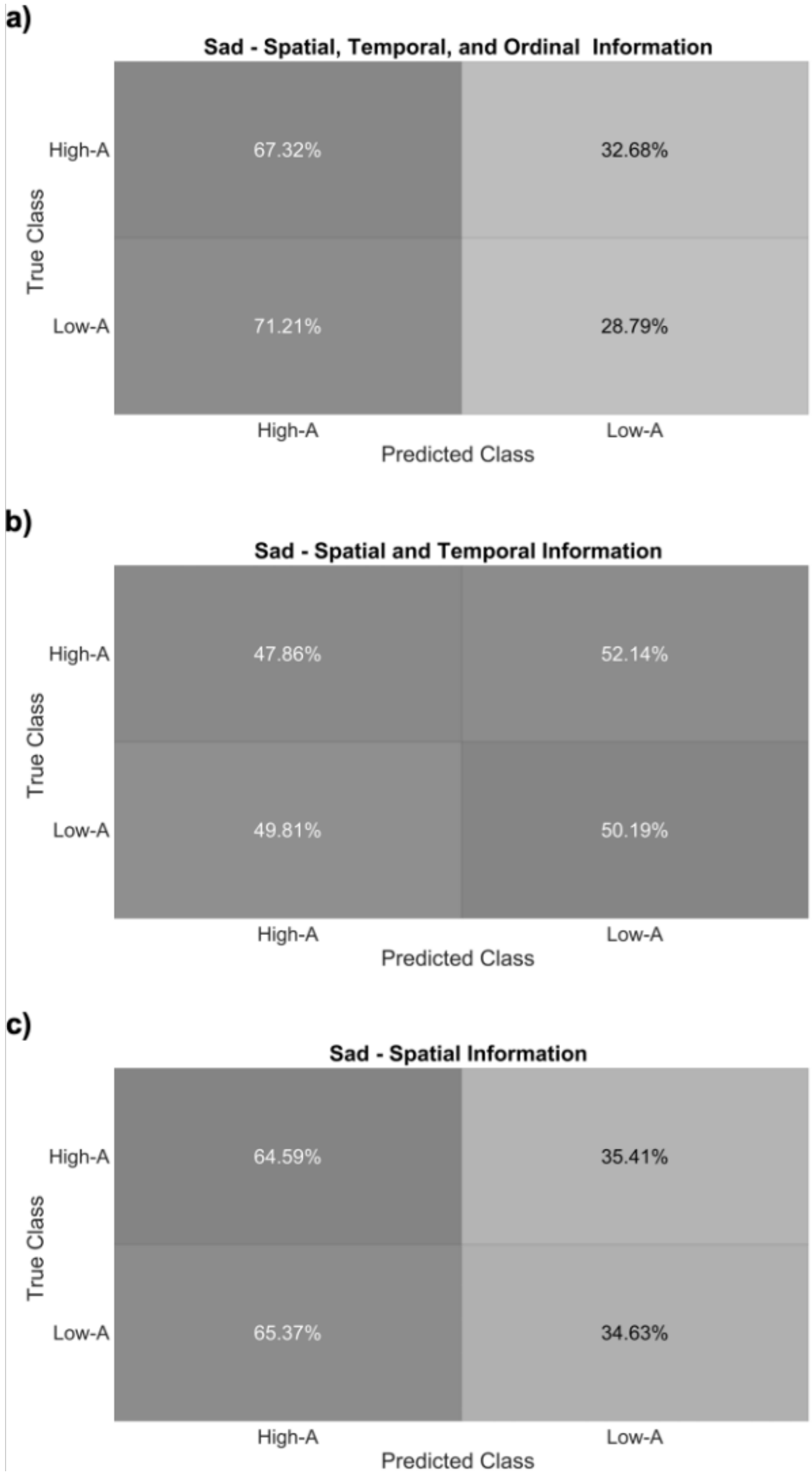


Figure 22. Confusion matrices showing classification accuracy and confusion percentages for all sad classification conditions when the classifier was trained on ASD

High-A/ Low-A data and tested on NT High-A/ Low-A data. (a) Classification accuracy when using Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

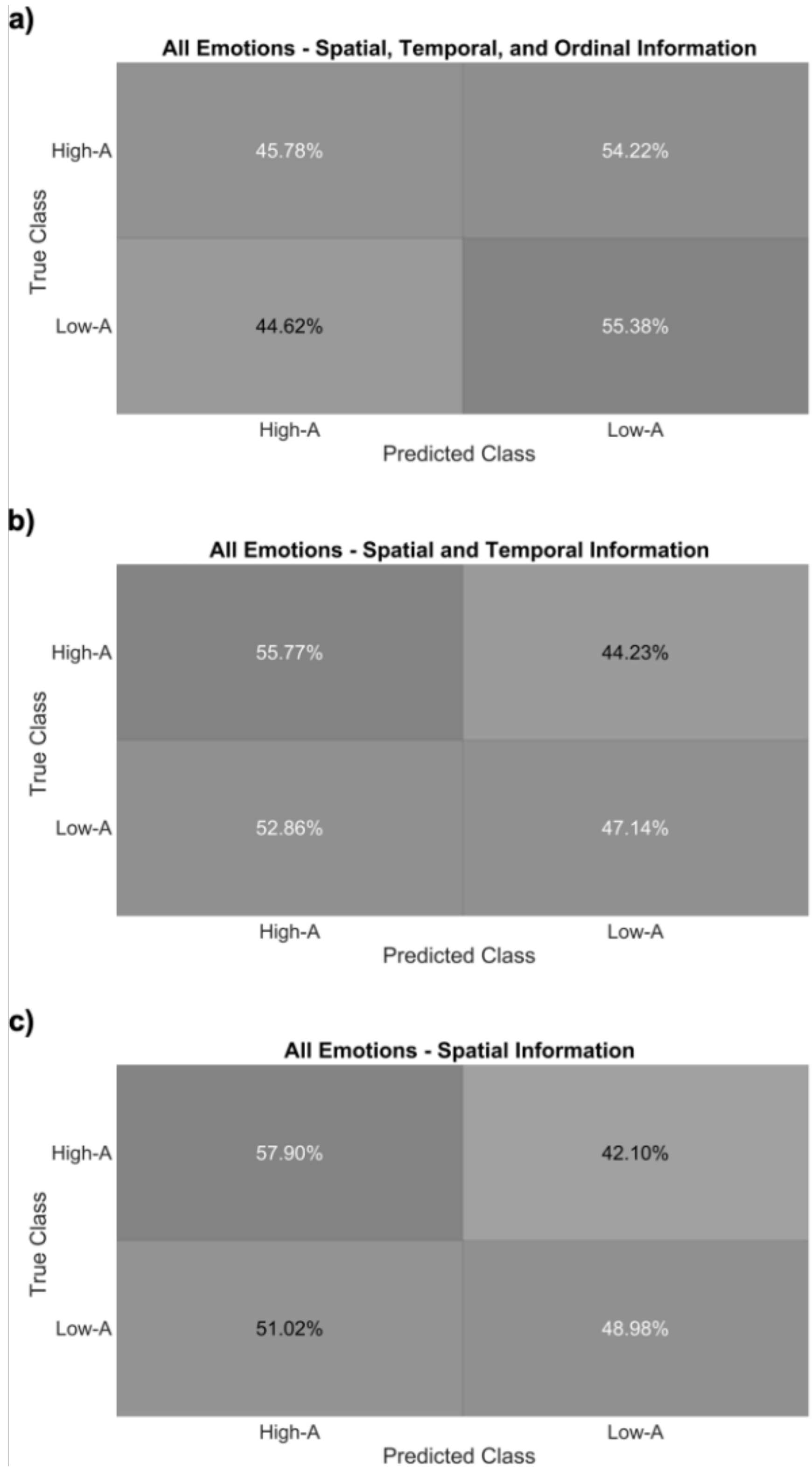


Figure 23. Confusion matrices showing classification accuracy and confusion percentages for all emotional expressions classification conditions when the classifier

was trained on ASD High-A/ Low-A data and tested on NT High-A/ Low-A data. (a) Classification accuracy when using Spatial-Temporal-Ordinal scanpath images. (b) Classification accuracy when using Spatial-Temporal scanpath images. (c) Classification accuracy when using Spatial scanpath images.

Table 1

Decoding Accuracy for ASD High-A vs. ASD Low-A vs. NT High-A vs. NT Low-A

Scanpath Information	Emotional Expression Information				
	Angry	Fear	Happy	Sad	All Expressions
Spatial-Temporal-Ordinal	50.68%***	54.18%***	55.35%***	55.55%***	59.43%***
Spatial-Temporal	47.76%***	47.08%***	49.32%***	47.08%***	54.83%***
Spatial	43.58%***	43.58%***	47.96%***	46.98%***	52.30%***

Note. *** indicates p <.001

Table 2

Decoding Accuracy for High-A vs. Low-A (ASD Training Data, NT Testing Data)

Scanpath Information	Emotional Expression Information				
	Angry	Fear	Happy	Sad	All Expressions
Spatial-Temporal-Ordinal	45.91%	49.81%	54.67%***	48.05%	50.58%
Spatial-Temporal	49.22%	48.25%	54.09%***	49.03%	51.46%
Spatial	54.67%***	55.64%***	54.48%***	49.61%	53.44%***

Note. *** indicates p <.001

APPENDICES

Appendix A

Autism Spectrum Quotient 10 (AQ-10) Questionnaire

NHS National Institute for Health Research

AQ-10

Autism Spectrum Quotient (AQ)

A quick referral guide for adults with suspected autism who do not have a learning disability.

	Please tick one option per question only:	**Definitely Agree**	**Slightly Agree**	**Slightly Disagree**	**Definitely Disagree**
1	I often notice small sounds when others do not				
2	I usually concentrate more on the whole picture, rather than the small details				
3	I find it easy to do more than one thing at once				
4	If there is an interruption, I can switch back to what I was doing very quickly				
5	I find it easy to 'read between the lines' when someone is talking to me				
6	I know how to tell if someone listening to me is getting bored				
7	When I'm reading a story I find it difficult to work out the characters' intentions				
8	I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant etc)				
9	I find it easy to work out what someone is thinking or feeling just by looking at their face				
10	I find it difficult to work out people's intentions				

SCORING: Only 1 point can be scored for each question. *Score 1 point for Definitely or Slightly Agree on each of items 1, 7, 8, and 10. Score 1 point for Definitely or Slightly Disagree on each of items 2, 3, 4, 5, 6, and 9.* If the individual scores **more than 6 out of 10**, consider referring them for a specialist diagnostic assessment.

This test is recommended in 'Autism: recognition, referral, diagnosis and management of adults on the autism spectrum' (NICE clinical guideline CG142). www.nice.org.uk/CG142

Key reference: Allison C, Auyeung B, and Baron-Cohen S, (2012) *Journal of the American Academy of Child and Adolescent Psychiatry* 51(2):202-12.

UNIVERSITY OF CAMBRIDGE

autism research centre arc

Appendix B

Toronto Alexithymia Scale (TAS-20) Questionnaire

Toronto Alexithymia Scale (TAS) Assessment

Client Information

Name:

Date:

Instructions

Please read each statement carefully and indicate how much you agree or disagree with the statement. Rate your level of agreement on a scale of 1 to 5, where:

- **1** represents **"Strongly Disagree"**
- **2** represents **"Disagree"**
- **3** represents **"Neither Agree nor Disagree"**
- **4** represents **"Agree"**
- **5** represents **"Strongly Agree"**

Assessment

Statement	Rating (1-5)
1. I am often confused about what emotion I am feeling.	
2. It is difficult for me to find the right words for my feelings.	
3. I have physical sensations that even doctors don't understand.	
4. I am able to describe my feelings easily.	
5. I prefer to analyze problems rather than just describe them.	
6. I prefer talking to people about their daily activities rather than their feelings.	

7. I find it hard to imagine what it would be like to be in someone else's shoes.	
8. When I am upset, I don't know if I am sad, frightened, or angry.	
9. I am often puzzled by sensations in my body.	
10. I prefer to just let things happen rather than understand why they turned out that way.	
11. I have feelings that I can't quite identify.	
12. Being in touch with emotions is essential.	
13. I find it hard to describe how I feel about people.	
14. People tell me to describe my feelings more.	
15. I don't know what's going on inside me.	
16. I often don't know why I am angry.	
17. I prefer talking to people about their daily activities rather than their feelings.	
18. I prefer to watch "light" entertainment shows rather than psychological dramas.	
19. It is difficult for me to reveal my innermost feelings, even to close friends.	
20. I can feel close to someone, even in moments of silence.	
Total Score:	

Practitioner's Signature: ________________________

Thank you for completing this assessment. Your responses will help us to better understand your emotional processing and guide our treatment approach.

Toronto Alexithymia Scale (TAS) Interpretation

Client Information

Name:

Date of Assessment:

Total TAS Score: _____

Interpretation:

The Toronto Alexithymia Scale (TAS) measures the degree of alexithymia, which refers to difficulties in identifying and describing emotions. The scale ranges from 20 to 100, with higher scores indicating higher levels of alexithymia.

- **Score below 52:** No alexithymia. Client demonstrates a healthy ability to recognize and describe emotions.
- **Score between 52 and 60:** Possible alexithymia. Client may have some difficulties recognizing or describing emotions. Further assessment may be necessary.
- **Score above 60:** High alexithymia. Client has significant difficulties recognizing or describing emotions. Therapeutic interventions should focus on improving emotional awareness and expression.

Recommendations:

Follow-up Actions:

Practitioner's Signature: __________________

REFERENCES

Adolph, D., & Alpers, G. W. (2010). Valence and arousal: a comparison of two sets of emotional facial expressions. *The American Journal of Psychology*, *123*(2), 209-219.

Adolphs, R. (1999) Social cognition and the human brain. *Trends in Cognitive Sciences, 3*(12), 469–479.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669-672.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*(6684), 470-474.

Ahmed, I. A., Senan, E. M., Rassem, T. H., Ali, M. A., Shatnawi, H. S. A., Alwazer, S. M., & Alshahrani, M. (2022). Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques. Electronics, 11(4), 530.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Bae, G. Y., & Luck, S. J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *Journal of Neuroscience*, *38*(2), 409-422.

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development*, 41–59.

Benton, C. P., Etchells, P. J., Porter, G., Clark, A. P., Penton-Voak, I. S., & Nikolov, S. G. (2007). Turning the other cheek: The viewpoint dependence of facial expression after-effects. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1622), 2131–2137. doi:10. 1098/rspb.2007.0473

Ben-Shalom, D., Mostofsky, S. H., Hazlett, R. L., Goldberg, M. C., Landa, R. J., Faran, Y., et al. (2006). Normal physiological emotions but differences in expression of conscious feelings in children with high-functioning autism. Journal of Autism and Developmental Disorders, 36(3), 395–400.

Bi, X., Guo, M., Cao, J., & Hao, Y. (2022). The study of face processing in social anxiety disorder based on face-specific N170 component. *Journal of Healthcare Engineering*, *2022*.

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Translational psychiatry*, *3*(7), e285-e285.

Bird, G., Press, C., & Richardson, D. C. (2011). The role of alexithymia in reduced eye-fixation in autism spectrum conditions. *Journal of autism and developmental disorders*, *41*(11), 1556-1564.

Black, M. H., Chen, N. T., Iyer, K. K., Lipp, O. V., Bölte, S., Falkmer, M., Tan, T., Girdler, S. (2017). Mechanisms of facial emotion recognition in autism spectrum disorders: Insights from eye tracking and electroencephalography. *Neuroscience & Biobehavioral Reviews*, *80*, 488-515.

Blakemore, C., & Sutton, P. (1969). Size adaptation: A new aftereffect. *Science, 166*(3902), 245-247

Blair, R. J. R. (1999). Psychophysiological responsiveness to the distress of others in children with autism. *Personality and Individual Differences, 26*(3), 477–485.

Bothe, E., Palermo, R., Rhodes, G., Burton, N., & Jeffery, L. (2019). Expression recognition difficulty is associated with social but not attention-to-detail autistic traits and reflects both alexithymia and perceptual difficulty. *Journal of Autism and Developmental Disorders*, *49*(11), 4559-4571.

Bruce, V. (1979). Searching for politicians: An information-processing approach to face recognition. *Quarterly Journal of Experimental Psychology*, *31*(3), 373-396.

Bruce, V. (1983). Recognizing faces. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *302*(1110), 423-436.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, *77*(3), 305-327.

Burton, N., Jeffery, L., Skinner, A. L., Benton, C. P., & Rhodes, G. (2013). Nine-year-old children use norm-based coding to visually represent facial expression. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1261.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49-62.

Cuve, H. C., Castiello, S., Shiferaw, B., Ichijo, E., Catmur, C., & Bird, G. (2021). Alexithymia explains atypical spatiotemporal dynamics of eye gaze in autism. *Cognition*, *212*, 104710.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., & Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature neuroscience*, *8*(4), 519-526.

Darwin, C. (1872). The Expression of the Emotions in Man and Animals. London: J. Murray.

Damasio, A. R., Tranel, D., and Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual Review Neuroscience, 13*: 89–109.

Desai, A., Foss-Feig, J. H., Naples, A. J., Coffman, M., Trevisan, D. A., & McPartland, J. C. (2019). Autistic and alexithymic traits modulate distinct aspects of face perception. *Brain and cognition*, *137*, 103616.

Dong, Z., Wang, G., Lu, S., Li, J., Yan, W., & Wang, S. J. (2022). Spontaneous facial expressions and micro-expressions coding: From brain to face. *Frontiers in Psychology*, *12*, 784834.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual review of vision science*, *1*, 393-416.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124.

Evers, K., Steyaert, J., Noens, I., & Wagemans, J. (2015). Reduced recognition of dynamic facial emotional expressions and emotion-specific response bias in children with an autism spectrum disorder. *Journal of autism and developmental disorders*, *45*, 1774-1784.

Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias: Implications for visual object representation. *Human Cognitive Neuropsychology.*, *8*: 1–19.

Farajzadeh, N., & Hashemzadeh, M. (2018). Exemplar-based facial expression recognition. *Information Sciences*, *460*, 318-330.

Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Bölte, S., Baron-Cohen, S., & Golan, O. (2016). Basic and complex emotion recognition in children with autism: cross-cultural findings. *Molecular autism*, *7*(1), 1-11.

Fujiwara, E. (2018). Looking at the eyes interferes with facial emotion recognition in alexithymia. *Journal of Abnormal Psychology*, *127*(6), 571.

Garcia, J. O., Srinivasan, R., & Serences, J. T. (2013). Near-real-time feature-selective modulations in human cortex. *Current Biology*, *23*(6), 515-522.

Gorno-Tempini, M. L., Pradelli, S., Serafini, M., Pagnoni, G., Baraldi, P., Porro, C., & Nichelli, P. (2001). Explicit and incidental facial expression processing: an fMRI study. *Neuroimage*, *14*(2), 465-473.

Gorno-Tempini, M. L., Price, C. J., Josephs, O., Vandenberghe, R., Cappa, S. F., Kapur, N., ... & Tempini, M. L. (1998). The neural systems sustaining face and proper-name processing. *Brain: a journal of neurology*, *121*(11), 2103-2118.

Greening, S. G., Mitchell, D. G., & Smith, F. W. (2018). Spatially generalizable representations of facial expressions: Decoding across partial face samples. *Cortex*, *101*, 31-43.

Griffiths, S., Jarrold, C., Penton-Voak, I. S., Woods, A. T., Skinner, A. L., & Munafò, M. R. (2019). Impaired recognition of basic emotions from facial expressions in young people with autism spectrum disorder: Assessing the importance of expression intensity. *Journal of autism and developmental disorders*, *49*(7), 2768-2778.

Grootswagers, T., Wardle, S.G., Carlson, T.A., 2017. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of Cognitive Neuroscience., 29* (4), 677–697. doi: 10.1162/jocn_a_01068.

Gwinn, O. S., Matera, C. N., O'Neil, S. F., & Webster, M. A. (2018). Asymmetric neural responses for facial expressions and anti-expressions. *Neuropsychologia*, *119*, 405-416.

Hanley, M., McPhillips, M., Mulhern, G., & Riby, D. M. (2013). Spontaneous attention to faces in Asperger syndrome using ecologically valid static stimuli. *Autism*, *17*(6), 754-761.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychology review*, *20*, 290-322.

Hasselmo, M. E., Rolls, E. T., and Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioral Brain Research*, *32*: 203–218.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*(6), 223-233.

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron*, *22*(1), 189-199.

Hedger, N., Adams, W. J., & Garner, M. (2015). Fearful faces have a sensory advantage in the competition for awareness. *Journal of experimental psychology: human perception and performance*, *41*(6), 1748.

Hileman, C. M., Henderson, H., Mundy, P., Newell, L., & Jaime, M. (2011). Developmental and individual differences on the P1 and N170 ERP components in children with and without autism. *Developmental neuropsychology*, *36*(2), 214-236.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature neuroscience*, *3*(1), 80-84.

Hong, S. W., & Yoon, K. L. (2018). Intensity dependence in high-level facial expression adaptation aftereffect. *Psychonomic Bulletin & Review*, *25*(3), 1035–1042. doi: 10.3758/s13423-017-1336-2

Jeffery, L., McKone, E., Haynes, R., Firth, E., Pellicano, E., & Rhodes, G. (2010). Four-to-six-year-old children use norm-based coding in face-space. *Journal of Vision*, *10*(5), 18–18. doi:10.1167/10.5.18

Jeffery, L., Rhodes, G., McKone, E., Pellicano, E., Crookes, K., & Taylor, E. (2011). Distinguishing norm-based from exemplar-based coding of identity in children: Evidence from face identity aftereffects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1824.

Jiang, M., Francis, S. M., Srishyla, D., Conelea, C., Zhao, Q., & Jacob, S. (2019, July). Classifying individuals with ASD through facial emotion recognition and eye-tracking. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6063-6068). IEEE.

Kang, J., Han, X., Song, J., Niu, Z., & Li, X. (2020). The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. *Computers in biology and medicine*, *120*, 103722.

Kang, E., Keifer, C. M., Levy, E. J., Foss-Feig, J. H., McPartland, J. C., & Lerner, M. D. (2018). Atypicality of the N170 event-related potential in autism spectrum disorder: a meta-analysis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(8), 657-666.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience, 17*(11), 4302-4311.

Kessels, R., Spee, P., & Hendriks, A. (2010). Perception of dynamic facial emotional expressions in adolescents with autism spectrum disorders (ASD). *Translational Neuroscience*, *1*(3), 228-232.

Kliemann, D., Dziobek, I., Hatri, A., Baudewig, J., & Heekeren, H. R. (2012). The role of the amygdala in atypical gaze on emotional faces in autism spectrum disorders. *Journal of Neuroscience*, *32*(28), 9469-9476.

Lane, R. D., Lee, S., Reidel, R., Weldon, V., Kaszniak, A., & Schwartz, G. E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomatic medicine*, *58*(3), 203-210.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Technical manual and affective ratings Gainesville: University of Florida, Center for Research in Psychophysiology.

Leopold, D. A., & Bondar, I. (2005). Adaptation to complex visual patterns in humans and monkeys. *Fitting the mind to the world: Adaptation and after-effects in high-level vision*, 189-211.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*,572–575. doi:10.1038/nature04951

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature neuroscience*, *4*(1), 89-94.

Leung, R. C., Pang, E. W., Anagnostou, E., & Taylor, M. J. (2018). Young adults with autism spectrum disorder show early atypical neural activity during emotional face processing. *Frontiers in human neuroscience*, *12*, 57.

Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *Journal of Neuroscience*, *20*(2), 878-886.

Liang, Y., Liu, B., Xu, J., Zhang, G., Li, X., Wang, P., & Wang, B. (2017). Decoding facial expressions based on face-selective and motion-sensitive areas. *Human brain mapping*, *38*(6), 3113-3125.

Lithfous, S., & Rossion, B. (2018). Electrophysiological individual face adaptation effects with fast periodic visual stimulation resist long interruptions in adaptation. *Biological psychology*, *133*, 4-9.

Liu, X., Liao, Y., Zhou, L., Sun, G., Li, M., & Zhao, L. (2013). Mapping the time course of the positive classification advantage: an ERP study. *Cognitive, Affective, & Behavioral Neuroscience*, *13*, 491-500.

Lu, S., Li, J., Wang, Y., Dong, Z., Wang, S. J., & Fu, X. (2022). A more objective quantification of micro-expression intensity through facial electromyography. In *Proceedings of the 2nd Workshop on Facial Micro-Expression: Advanced Techniques for Multi-Modal Facial Expression Analysis* (pp. 11-17).

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institute, ISBN 91-630-7164-9.

Mares, I., Ewing, L., Farran, E. K., Smith, F. W., & Smith, M. L. (2020). Developmental changes in the processing of faces as revealed by EEG decoding. *Neuroimage*, *211*, 116660.

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach-and avoidance-related behaviors. *Emotion*, *5*(1), 119.

McCollough, C. (1965). Color adaptation of edge-detectors in the human visual system. *Science*, *149*(3688), 1115-1116.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Muckli, L., De Martino, F., Vizioli, L., Petro, L.S., Smith, F.W., Urgubil, K., Goebel, R., Yacoub, E., 2015. Contextual Feedback to Superficial Layers of V1. *Current Biology, 25*, 2690–2695.

Nemrodov, D., Ling, S., Nudnou, I., Roberts, T., Cant, J. S., Lee, A. C., & Nestor, A. (2019). A multivariate investigation of visual word, face, and ensemble processing: Perspectives from EEG-based decoding and feature selection. *Psychophysiology*, e13511.

Neth, D., & Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *Journal of vision*, *9*(1), 5-5.

Nuske, H. J., Vivanti, G., & Dissanayake, C. (2014). Reactivity to fearful expressions of familiar and unfamiliar people in children with autism: an eye-tracking pupillometry study. *Journal of Neurodevelopmental Disorders*, *6*(1), 1-16.

Nuske, H. J., Vivanti, G., Hudry, K., & Dissanayake, C. (2014). Pupillometry reveals reduced unconscious emotional reactivity in autism. *Biological psychology*, *101*, 24-35.

Ojemann, J. G., Ojemann, G. A., and Lettich, E. (1992). Neural activity related to faces and matching in human right nondominant temporal cortex. *Brain*, *115*: 1–13.

Ola, L., & Gullon-Scott, F. (2020). Facial emotion recognition in autistic adult females correlates with alexithymia, not autism. *Autism*, *24*(8), 2021-2034.

Petro, L.S., Smith, F.W., Schyns, P.G., Muckli, L., 2013. Decoding face categories in diagnostic subregions of primary visual cortex. *European Journal of Neuroscience*, *37*, 1130–1139.

Posamentier, M. T., & Hervé A. (2003) Processing faces and facial expressions. *Neuropsychology review, 13*(3), 113-143.

Posserud, M., Lundervold, A. J., Lie, S. A., & Gillberg, C. (2010). The prevalence of autism spectrum disorders: impact of diagnostic instrument and non-response bias. *Social psychiatry and psychiatric epidemiology*, *45*, 319-327.

Rhodes, G. (1985). Lateralized processes in face recognition. *British journal of Psychology*, *76*(2), 249-271.

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, *46*, 2977–2987.

Robbins, R., McKone, E., & Edwards, M. (2007). Aftereffects for face attributes with different natural variability: Adapter position effects and neural models. *Journal of Experimental Psychology: Human Perception and Performance*, *33*,570–592. doi:10.1037/0096-1523.33.3.570

Ross, D. A., Deroche, M., & Palmeri, T. J. (2014). Not just the norm: Exemplar-based models also predict face aftereffects. *Psychonomic Bulletin & Review*, *21*(1), 47-70.

Sawada, R., & Sato, W. (2015). Emotional attention capture by facial expressions. *Scientific reports*, *5*(1), 1-8.

Sawada, R., Sato, W., Uono, S., Kochiyama, T., & Toichi, M. (2014). Electrophysiological correlates of the efficient detection of emotional facial expressions. *Brain research*, *1560*, 60-72.

Skinner, A. L., & Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial expressions. *Psychological Science*, 21(9), 1248-1253.

Smith, F.W., Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proc. Natl. Acad. Sci. U. S. A., 107*, 20099–20103.

Smith, F. W., & Smith, M. L. (2019). Decoding the dynamic representation of facial expressions of emotion in explicit and incidental tasks. *Neuroimage*, *195*, 261-271.

Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. *Neuroimage*, *35*(2), 836-852.

Sprengelmeyer, R., Rausch, M., Eysel, U. T., & Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1409), 1927-1931.

Streit, M., Ioannides, A. A., Liu, L., Wölwer, W., Dammers, J., Gross, J., & Müller-Gärtner, H. W. (1999). Neurophysiological correlates of the recognition of facial expressions of emotion as revealed by magnetoencephalography. *Cognitive brain research*, *7*(4), 481-491.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual review of psychology, 63,* 483-509.

Tottenham, N., Hertzig, M. E., Gillespie-Lynch, K., Gilhooly, T., Millner, A. J., & Casey, B. J. (2014). Elevated amygdala response to faces and gaze aversion in autism spectrum disorder. *Social cognitive and affective neuroscience, 9*(1), 106-117.

Tregillus, K. E., & Engel, S. A. (2019). Long-term adaptation to color. *Current Opinion in Behavioral Sciences*, *30*, 116-121.

Trevisan, D. A., Hoskyn, M., & Birmingham, E. (2018). Facial expression production in autism: A meta-analysis. *Autism Research*, *11*(12), 1586-1601.

Tseng, A., Bansal, R., Liu, J., Gerber, A. J., Goh, S., Posner, J., ... & Peterson, B. S. (2014). Using the circumplex model of affect to study valence and arousal ratings of emotional faces by children and adults with autism spectrum disorders. *Journal of autism and developmental disorders*, *44*(6), 1332-1346.

Van der Donck, S., Dzhelyova, M., Vettori, S., Thielen, H., Steyaert, J., Rossion, B., & Boets, B. (2019). Fast periodic visual stimulation EEG reveals reduced neural sensitivity to fearful faces in children with autism. *Journal of Autism and Developmental Disorders*, *49*(11), 4658-4673.

Waller, J. T., & Svensson, E. I. (2017). Body size evolution in an old insect order: no evidence for Cope's Rule in spite of fitness benefits of larger size. *Evolution*, *71*(9), 2178-2193.

Wang, J., & Perez, L. (2017). The effectiveness of data augmentation in image classification using deep learning. Convolutional Neural Networks Vis. Recognit, 11(2017), 1-8.

Wang, S., Tudusciuc, O., Mamelak, A. N., Ross, I. B., Adolphs, R., & Rutishauser, U. (2014). Neurons in the human amygdala selective for perceived emotion. *Proceedings of the National Academy of Sciences*, *111*(30), E3110-E3119.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*(6982), 557-561.

White, S. W., Maddox, B. B., & Panneton, R. K. (2015). Fear of negative evaluation influences eye gaze in adolescents with autism spectrum disorder: A pilot study. *Journal of Autism and Developmental Disorders*, *45*(11), 3446-3457.

Xu, Y., Jia, R., Mou, L., Li, G., Chen, Y., Lu, Y., & Jin, Z. (2016). Improved relation classification by deep recurrent neural networks with data augmentation. arXiv preprint arXiv:1601.03651.

Yang, H., Susilo, T., & Duchaine, B. (2016). The anterior temporal face area contains invariant representations of face identity that can persist despite the loss of right FFA and OFA. *Cerebral Cortex*, *26*(3), 1096-1107.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, *81*(1), 141.