Spring 5-1-2022

Autistic Individuals May See Emotions Differently: How Gradable Adjectives Can Be Used to Determine Emotion Recognition

Devon Murphy
devon.murphy@uconn.edu

Autistic Individuals May See Emotions Differently: How Gradable Adjectives Can Be Used to Determine Emotion Recognition

Devon Murphy

University of Connecticut

Department of Psychological Sciences

Thesis Supervisor: Dr. Letitia Naigles

Honors Advisor: Dr. James Chrobak

Apr 29, 2022

Abstract

Autism spectrum disorder is characterized by multiple differences in functioning, with a strong correlation to perceived language deficits. Evidence suggests that during childhood, children with ASD lag behind their TD peers in terms of emotion recognition skills. However, the prior research had not pushed into adolescence in studying these phenomena. The current study seeks to fill in that gap, and look at autistic individuals in adolescence to determine whether the previous research holds up as they age, or whether there is a significant change in skills. This study is a part of a longitudinal study on language acquisition in autistic individuals, which follows participants from early in childhood through adolescence and young adulthood. The current study looks at 26 participants (TD and ASD), each of whom were visited on two separate occasions, approximately two weeks apart, in order to collect data. Part of the data collected was for the emotion recognition portion of the study, and was examining in particular the use of gradable adjectives and perspective-taking. The results indicate that these participants have caught up to their TD peers. Findings are discussed in terms of increasing understanding of language acquisition in autistic individuals, particularly as they age.

Keywords: autism, gradable adjectives, adolescence, emotion recognition

Introduction

The ability for perspective taking in our world is incredibly important. For most of us, looking at the context surrounding something is intuitive and natural. For example, when looking at a group of kindergarteners, the tallest of the group would be labeled ‘tall’. However, would the same be said if you placed that ‘tall’ kindergartener with a group of adult basketball players? For some, the answer is clearly no, when surrounded by basketball players, the kindergartener would no longer be classified as tall. For others, this classification may not be so easy. We seek to better understand these perceived challenges in autistic individuals through the use of gradable adjectives in our study. Gradable adjectives are adjectives that can vary in intensity and depend on the speaker's perspective. Adjectives can be relative or absolute, and can be physical and emotional. Some examples of absolute physical adjectives are spotted and striped. No matter how many stripes are on an umbrella, if it is striped, it should be classified as striped, even if there are other umbrellas with more stripes. In other words, there is a concrete, correct answer to the question, “How many striped umbrellas are there?” However, gradable adjectives are different. For the physical dimension, an example would be ‘long’. This is relative because the context of the stimuli may change the participants ideas of what constitutes long. If there are a hundred 1 inch long straws and one 5 inch long straw, most people would consider the five inch long straw to be long. However, if there were several five inch long straws and no one inch long straws, the one five inch long straw might not be considered long anymore. The same holds true for emotional dimensions (ex. Happy faces). Below, we will describe how these adjectives are acquired in typical development and how they seem to be problematic for autistic individuals. The current study builds on the previous research by looking at autistic individuals in adolescence, where most of the previous studies have focused on younger children with autism.

Previous studies about gradable adjectives

Subjectivity in language is important, and relates strongly to the current study because disagreement over what is correct in terms of categorization is possible, and indeed, probable. Subjectivity is the idea that people may see things slightly differently, and that does not necessarily make either party wrong. For example, two people can look at a statue, and one could determine that it is tall, and the other that it is medium height, based on their perspective and the context they are comparing it to. Adults are clearly able to understand this subjectivity, and permit disagreement within reason. For example, Foushee et al. (2017) introduced adults to distinct sets of exemplars belonging to the same novel kind, labeling them pimwits or daxes. Pimwits were thin purple cylinders ranging from 0.75 to 6.25 inches in height, and from densely to sparsely spotted. Daxes were blue and yellow spheres ranging from 0.5 to 3 inches in diameter, and from heavily to lightly striped. In the experimental narrative, two animated characters were described as seeing different exemplars, for example one character saw five relatively short and densely spotted pimwits, while the second character saw five relatively tall and lightly spotted exemplars. Then the characters would disagree with each other on, for example, whether the pimwits were tall, and participants were asked on each trial whether each speaker 'could be right' or not, and why. Foushee et al. found that disagreements over relative, subjective GAs (e.g. tall) were almost always judged as 'faultless'; that is, that both speakers could be right. Whereas those over absolute (e.g. spotted, when both had spots) adjectives were more likely to have just one correct speaker. When 9-year-old children were tested, their judgements were similar to those of adults, permitting faultless disagreement much more often for 'tall' than for 'spotted'. Younger children, though, generated mixed results. This provides evidence for consideration of speaker at the level of semantics, and adults' sensitivity to the potential for differing standards of more than just explicitly context-dependent adjectives. While this study did not intersect with autism in particular, it does provide information pertaining to the effect of context on sorting practices.

An interesting study looking at the particulars of gradable adjectives, and which was very helpful in the framing of the ideas of the current study was Britton et al. (2017). Here, the researchers focused directly on the distinction between absolute and gradable adjectives, and how absolute adjectives are independent of seeding. Seeding is when you add more "striped" umbrellas, for example, to see whether that changes if the participant chooses *all* the "striped" umbrellas to put in the box. There were two conditions, first with the nine basic stimuli, and then with an additional four instances of the most extreme physical expression added. An example of this setup is having nine pictures of striped umbrellas and unstriped umbrellas, where the participant is supposed to decide which of the pictures are of a striped umbrella. The second part of the experiment is when there are four more pictures showing unstriped umbrellas. This is to see whether adding more unstriped umbrellas will confuse the participants as to how many striped umbrellas there are. The experiment done on children (ages 4-7 years old) showed that for the striped umbrellas, there was a zero mean shift, meaning having more unstriped umbrellas did not change which umbrellas the participants perceived as striped because it was an absolute adjective. This was important given the way we chose our control variables, because absolute adjectives should be independent of seeding, meaning no matter how many non-striped umbrellas there are, that doesn't change the number of striped umbrellas.

Another study looked at the distinction between relative and absolute adjectives with respect to impossible/possible requests. In Syrett et al. (2006) the participants were first exposed to the Presupposition Assessment Task, where they were presented with a series of pairs of objects accompanied by a request, and the context either satisfied or violated the presuppositions of a singular definite NP. Participants were expected to accept a felicitous request and reject an infelicitous request (e.g. which is the curved line?, when asked to compare a straight line and a curved line versus two curved lines). The researchers looked at the children's responses, as well as the time it took for them to form their responses. There were two different absolute adjectives (straight and bumpy), and two sets of straight stimuli, with 2 orders each (early, late). There were 4 target adjectives: relative (big, long) and absolute (full, spotted). There were 2 items per target adjective (big/small, long/short, one/neither full, one/both spotted), in addition to 3 orders (early, early 2, late) based on location of 'non-full/non-full' pair: early in sequence (before or after relative pair), or later. The results showed that children and adults patterned similarly in their acceptance and rejection of the request, namely that they accepted the felicitous request faster than the non-felicious request, with respect to relative GAs and minimum absolute GAs, with one exception, the full/straight pairs. Children were more likely to accept the non-maximal member as 'the full/straight one'. There was also an order effect for both the full and straight pairs, where children were more correct when the maximal pair (and no relative pair) preceded the non-maximal pair. Children also took significantly longer to identify the non-full object as the full one than the full object, given the same request. The children seemed to be more willing to accept uses that are 'close enough to true' in contexts where a precise use has not yet been exemplified. This study provided experimental evidence for distinction between semantic vagueness and pragmatic imprecision.

These studies, which have primarily focused on children, have introduced the ideas of the importance of using gradable adjectives to help determine the emotion recognition of autistic individuals. These studies have shown that children are less likely to make concessions, and think in more black and white terms, whereas adults are more comfortable accepting that disagreements can happen, and both parties can be correct. We use this information to better inform our own study, where we look at adolescents on the autism spectrum.

Autism Spectrum Disorder

Autism spectrum disorder is a developmental disability characterized by differences in behavior, communication, learning, and interactions (Diagnostic, 2022). For the purposes of this study, and the background research wherein, we are focused on the communication differences that are present in autism spectrum disorder. While a language disorder itself is not a diagnostic criterion for autism spectrum disorder (ASD), many on the spectrum struggle with various aspects of language, and research into language and autism is important for increased understanding in the field (American, 2013). The theories about language in autistic individuals range from typical but delayed development, to deviance in language development. Autistic individuals appear to demonstrate similar patterns and processes of grammatical development as language-matched TD children. However, pragmatic language use is particularly difficult for autistic individuals; for example, autistic individuals are not as good at narrative storytelling when compared to TD individuals. An example of this is that autistic children and adolescents tend to produce narratives that are harder to follow and less engaging than the TD controls. This relates to perspective-taking because narrative storytelling is about explaining something to another person, and to do so, one must be able to take the other's perspective in order to adequately relay the story (Naigles & Chin, 2015).

An important contribution to the field comes from Naigles and Tek (2017), who posit the thesis that 'form is easy, meaning is hard'. This is the idea that for autistic individuals, vocabulary growth is not actually that difficult, but rather it is building lexical/semantic organization that is hard. This means that the words themselves are not hard, but rather knowing how to connect them in one's representation is what is hard. For example, an autistic individual may understand the words they are using, but not how to shape them into sentences with meaning. An example is the expression 'He kicked the bucket', and the ensuing understanding, in context, of what that means. There are multiple possible meanings for this expression, depending on context. When talking about an old man, this expression may mean the man has died. However, when talking about a toddler, this expression may mean the toddler has literally kicked over a bucket. This is a perspective-taking exercise because the participants need to understand what the researcher is asking in context. This is important from the perspective of interventions, and how to best help autistic individuals with language in the real world. In the real world, perspective-taking is important because we must understand what others are saying in context in order to correctly understand what is happening. This is important with respect to the current study because we are trying to determine whether autistic individuals can understand the context of a situation, and adjust accordingly. Context matters, and our study looks to see how autistic individuals use context in developing meaning. The use of gradable adjectives in our study allows us to investigate the understanding of emotion recognition in autistic individuals.

Autistic individuals additionally have difficulty with categorization. This is related to our study because we are asking our participants to categorize pictures based on emotional or physical dimensions. The participants have to sort the pictures in order for researchers to determine their emotion recognition. Ellawadi et al. (2016) investigated the categorization abilities of 6 year-old children with ASD and 5-year-old TD children. The children were part of a longitudinal study, and at the study onset the two groups were matched on language production. However, by the time of the present study, the TD individuals had more advanced language as measured by standardized tests.

The researchers compared the children's performance across trials with typical, somewhat typical, and atypical images of birds, cars, chairs, and cats. The typical trial showed a stimulus that would be immediately expected for the category, like a blue jay. The somewhat typical trial showed a stimulus that might not be immediately categorized, like a woodpecker. The atypical trial showed a stimulus that might be hard to categorize, like a penguin. Children were auditorily prompted with the prompt "Is this a ___?", and expected to press a green button for yes and a red button for no, where both accuracy and reaction time were recorded. Children in both groups were significantly more accurate when responding to typical versus atypical stimuli. However, the TD group demonstrated stable performance across the experiment whereas the ASD group was less accurate on the somewhat typical trials that were preceded by an atypical trial. Moreover, language but not nonverbal IQ was found to be a significant predictor of category structure for the TD group; in contrast, nonverbal IQ but not language was the significant predictor for the ASD group. This suggests that autistic individuals and TD individuals categorize stimuli differently. For TD individuals, language is more important to categorization, whereas for autistic individuals, nonverbal IQ is more important. This is important within our current study because we are using categorization methods to determine emotion recognition.

Wit et al. (2008) investigated whether children with ASD look at positive faces differently than negative faces. The children were five years of age, matched on age and cognitive ability with 5-year-old TD children. The children's gaze was recorded for both length of gaze as well as movement of the eyes. Wit et al. found that total looking time was significantly higher in the TD group (M = 7.25 s, S.E. = 0.36) than the ASD group (M = 5.12 s, S.E. = 0.67); however, proportions of looking only showed group differences for the mouth region. Moreover, children with ASD with higher (i.e., more severely affected) scores on a social/communicative assessment looked less at the mouth ($r = -.62$, $p < .05$, one-sided). Both groups looked longer at the eye region for negative emotions compared to positive emotions. This research provides evidence that looking time at the mouth relates to socio-communicative functioning and that children with ASD recognize positive versus negative emotions differently (i.e. focusing on the eye region for negative emotions). This is important for our current research because we are using both positive and negative emotions, and will be looking to see if there are differences between the adjectives used for these emotions.

Finally, de Villiers et al. (2019) utilized the same set of participants, and a similar method, to the current study. The participants (TD: M_{age}=10.5, SD= 2.15; ASD: M_{age}=12.3, SD=3.34) were shown 9 sets of 9 stimuli each, and asked to put the ‘happy’, ‘long’, etc, stimuli in a designated box. Later during the visit, the children were asked to do the task again, but this time there were four additional stimuli introduced, being of the maximal intensity. This was done to determine whether the cutoff value would change when the number of maximal intensity pictures changed. When asked to sort the pictures, de Villiers et al. found that both groups clearly distinguished the relative adjectives from the absolute adjectives. The children’s criterion for “long” pencils shifted significantly upwards with the set of 13 items (p=.005 for TD, p=.045 for ASD), but their cutoff was unchanged for the absolute adjectives. Thus, these autistic individuals did not have any impairment of their semantic knowledge of physical gradable adjectives. However, these autistic individuals had much more difficulty sorting the emotion faces, as they were unsystematic and mis-sorted frequently. For example, they made mistakes, and did not sort contiguously as often as the TD children. This means that the participants may have, for example, 3,4,6,7,8,9, and have missed the 5, making their sorting imperfect. The results provide further evidence of the difficulty experienced by autistic individuals in reading others' emotions (de Villiers et al.); however, these children were still mostly middle-school-aged, and we wondered whether their emotional understanding would continue to develop through adolescence. In future comparative studies between the two visits (visit 8: the de Villier study and visit 9: the current study), it would be helpful to recode the de Villier data with the lenient metric in order to more accurately compare the two data sets.

Current Study

The literature summarized above shows the current ideas concerning autistic individuals, and how they process language and recognize emotions and physical dimensions that are context dependent. The literature shows that young, autistic individuals seem to struggle more with emotion recognition, perspective-taking, and categorization. Children with ASD seem to lag behind their TD peers in almost every facet of emotion recognition, and they gaze at emotional faces differently. However, where the previous research is lacking is including participants who are adolescents, and that is what the current study looks to address. Using participants from the longitudinal study of early language (LSEL: Naigles & Fein 2017), we seek to extend what is currently known about language and emotion recognition by continuing to follow these participants and assessing how their recognition skills look during adolescence. While this study does not seek to compare the visit 8 and visit 9 data, we will determine how these participants behave at adolescence, while later studies will focus on a comparative analysis. Our main questions for this study were:

1) Do gradable adjectives reveal differences in perspective taking in autistic individuals at adolescence? and

2) Does the type of gradable adjective matter with respect to perspective taking?

Through the experimental design and analysis, we hope to answer these questions to help understand language acquisition and recognition for autistic individuals and to give other researchers a stepping off point for therapeutic interventions.

Methods:

Participants:

Twelve individuals with ASD (1 female) and fifteen TD individuals (5 females), all monolingual English speakers, participated in this study as a part of an ongoing longitudinal study on language development in children with autism spectrum disorder (Naigles & Fein 2017). One TD participant was excluded from the final dataset because he only participated in the first day of data collection, not the second day. Participants were administered the ADOS-G (Autism Diagnostic Observation Schedule-Generic) to confirm the diagnosis of ASD, and the CELF (Clinical Evaluation of Language Fundamentals - 5th Edition) and DAS (Differential Ability Scales - 2nd Edition) to provide demographic information (see Table 1) (Lord, 2000; Wiig, 2013; Elliot, 2007). The participants engaged in multiple games and tasks for the purposes of the longitudinal study, however, only the gradable adjective tasks were used for this analysis.

Stimuli:

There were nine sets of stimuli, including both emotion and physical dimension categories. For the emotion category, the stimuli included the faces of: happy man, happy woman, sad man, sad woman, angry man, angry woman. For these emotion stimuli, participants were asked to put the "happy", "angry", etc. faces in a designated box. There was also a physical dimension set, which included 9 pencils of varying lengths. For these stimuli, the task was the same, with participants being asked to put the "long" pencils in the box. Finally, there were two control sets of stimuli, striped umbrellas and spotted dogs, which are considered absolute adjectives, and therefore had a correct answer (i.e., when told to put the striped umbrellas in the box, all umbrellas with stripes were targets).

Each stimulus set had 9 pictures or pencils for visit 1. Then, for visit 2 (which occurred approximately two weeks after visit 1), the stimulus sets were increased by 4 of the maximal intensity pictures, for a total of 13 in each stimulus set. The participants were then asked to do the same activity, so that investigators could determine if the cutoff value changed between visits. The cutoff value is the lowest value the participant places in the box. It is the lowest value to which the participant assigns the given characteristic (i.e. "happy", "sad", "long", etc.).

Coding:

We recorded which of the stimuli the participants put in the box, and which they kept out of the box, and used this to determine the cutoff value for each adjective.

Cutoff values for the absolute adjectives were coded for their predetermined correct answers, as the dogs were either spotted or not spotted, and the umbrellas were either striped or not striped. There were five striped umbrellas, and five spotted dogs shown, and in order to be correct, the participants must have answered with the five spotted/striped pictures in each category. Therefore, if there were any mistakes in the sorting of these control stimuli, they were considered wrong.

Cutoff values for the relative adjectives were coded as:

For example: angry faces stimuli IN BOX: 4,5,6,7,8,9; angry faces stimuli OUT OF BOX: 1,2,3 for visit one, so the cutoff value would be 4 because that is the first stimulus in the 'angry' box, and therefore the lowest value emotion considered to be angry was '4'.

For visit two, the stimulus sets increased from nine pictures to thirteen pictures, and the cutoff values for the gradable adjectives were coded as:

For example: angry faces stimuli IN BOX: 5,6,7,8,9,9a,9b,9c,9d; angry faces stimuli OUT OF BOX: 1,2,3,4, so the cutoff value would be 5 because that is the first value in the 'angry' box; therefore, the lowest value emotion considered to be angry was '5'.

The cutoff values were calculated on a lenient basis, in that if there was only one mistake, that mistake was ignored (per advice from David Barner at UCSD, personal communication):

For example: angry faces stimuli IN BOX: 3,4,5,7,8,9; angry faces stimuli OUT OF BOX: 1,2,6, then 6 would be considered a mistake, and the cutoff value for faces to be considered angry would be 3.

Analysis Plan:

The goal of the analyses was to ascertain whether the increased number of stimuli in visit 2 changed the cutoff value, as well as whether there were any significant differences between the performance of the TD group vs. the ASD group. These data were analyzed using SPSS, specifically t-tests and general linear models. These analyses were done on both the lenient and non-lenient data. Lenient data afforded mistakes in sorting.

As described above, if the angry faces stimuli: 3,4,5,7,8,9 were all placed in the box, 3 would be considered the cutoff value, even though 6 was missing from the box. Non-lenient coding, on the other hand, afforded no such leniency in mistakes, and if the participants were not perfect in their sorting, their data was not used for each analysis. It is important to recognize that both the TD and ASD groups made mistakes, so the lenient data coding does not unfairly bias one group over the other. Only eight TD participants out of 15 made no mis-sorts. Two ASD participants out of 12 made no mis-sorts. The average number of mistakes by the TD group was 1.6, while the average number of mistakes by the ASD group was 3.25. The TD group did have slightly fewer mistakes, but the outliers affected the means as well.

For the results, the control stimuli were compared with a general linear model. Then, the average emotion cutoffs between visits 1 and 2 were measured with a t-test to determine if there was a significant difference by visit. Then each emotion (happy, sad, and angry) was analyzed individually with general linear models, in order to test for differences by visit. Then the man versus woman stimuli were first compared using a t-test, to determine whether the male stimuli yielded significantly different cutoff results than the female stimuli.

Finally, the shift data for each participant was analyzed using a chi-square test. Using only the emotion stimuli, we counted the number of times (i.e., number of stimuli) each participant shifted their cutoff score up from visit 1 to visit 2, the number of times they did not change their cutoff score between visits, and other patterns, in which the cutoff value changed in a different way between visit 1 and 2. If the participant shifted up at least 3/6 times, we coded that as generally going up. The same was true for staying the same, again at least 3 times. Otherwise, they were added to the 'other' group, in which there was no discernible pattern. We then did a chi-square analysis to determine whether these patterns differed between the TD group and the ASD group.

Results

Our main questions are whether the change in the number of stimuli will affect the cutoff value for specific emotion and physical attribute pictures using gradable adjectives, as well as whether there are significant differences between the TD and ASD groups.

We present analyses from both the lenient and non-lenient cutoffs. As described in the analysis plan, the lenient cutoffs offer a more comprehensive picture, due to bigger sample sizes. However, we also looked at the non-lenient cutoffs, to determine the effect of error in the lenient cutoffs. The non-lenient cutoffs had very small sample sizes, as any participant with a mistake had to be eliminated from the dataset.

Lenient coding:

As Table 3 shows, there does not appear to be any difference between visit 1 and visit 2, in either group, for the control stimuli. The mean cutoff for the spotted dogs was 5 for the TD and ASD groups in both visits, which showed that both groups understood the assignment, and there was no difference seen between the two groups for this stimulus set. The mean cutoff for the striped umbrellas was 4.75 (V1) and 4.83 (V2) for the ASD group and 5 for the TD group in both visits, which confirmed that both groups understood the assignment, with just a couple of ASD participants making mistakes in this control condition. This means that the participants understood the assignment, and that the experimental design was appropriate for their level.

Mean cutoffs for the pencils' stimuli can be seen in Table 3 for each group and visit. As the table shows, both groups increased their cutoff scores between visit 1 and visit 2 and the two groups do not appear to differ in this shifting. The GLM found a significant difference in cutoff scores between visit 1 and visit 2 (F=12.838, p=.001), no significant group differences (F=0.314, p=.580), and no significant interactions between group and visit (F=0.307, p=.584) (Figure 6).

Mean cutoffs for average emotions can be seen in Table 3 for each group and visit. As the table shows, there was a significant difference between visit 1 and 2 with respect to the cutoff values. The two-tailed paired t-test found a significant difference ($t(26)$=-3.077, p=0.005) in cutoff scores between visit 1 and visit 2, with significantly higher cutoffs for visit 2. However, the general linear model did not find significance between the groups (*F*=0.910, *p*=0.910) (Figure 7).

Mean cutoffs for angry faces can be seen in Table 3 for each group and visit. As the table shows, there was a significant difference between visit 1 and 2 with respect to the cutoff values, however, there does not appear to be a difference between the groups. The GLM found a significant difference in cutoff scores between visit 1 and visit 2 (*F*=14.805, *p*<.001), no significant group differences (*F*=0.335, *p*=.568), and no significant interactions between group and visit (*F*=2.280, *p*=.144) (Figure 8).

Mean cutoffs for sad faces can be seen in Table 3 for each group and visit. As the table shows, there did not seem to be any difference between visits or groups for sad faces. The GLM found no significant difference in cutoff scores between visit 1 and visit 2 (*F*=2.364, *p*=.137), no significant group differences (*F*=0.572, *p*=.457), and no significant interactions between group and visit (*F*=0.001, *p*=.975) (Figure 9).

Mean cutoffs for happy faces can be seen in Table 3 for each group and visit. As the table shows, there did not seem to be any difference between visits or groups for happy faces. The GLM found no significant difference in cutoff scores between visit 1 and visit 2 (*F*=3.445, *p*=.075), no significant group differences (*F*=1.048, *p*=.316), and no significant interactions between group and visit (*F*=0.319, *p*=.577) (Figure 10).

Mean cutoffs for man versus woman can be seen in Table 3 for each group and visit. As the table shows, there looks to be a difference in the cutoff values for female versus male faces. The two-sided paired t-test found a significant difference (*t*(26)=2.857, *p*=0.008) in cutoff scores between male and female stimuli, with the participants being more exclusive (i.e., including fewer instances of each emotion) for stimuli depicting men (Figure 11).

Non-lenient coding:

See Table 4 for the results of the non lenient data. As the table shows, there did not seem to be any difference between visits or groups. Analyses showed that for average emotions there were no significant differences, however, fewer participants could be included because of the prevalence of mistakes in sorting.

Table 4 shows the number of participants included in each non-lenient analysis, as well as the cutoff values and p-values from the analyses. Non-lenient analyses were performed on each individual emotion and physical dimension, as well as for collapsed emotions and gender analyses. Significance is denoted with an asterisk, where only the pencil dimension was significant between visits (there were no mistakes in this dimension, and therefore the analysis was the same for both lenient and non-lenient coding).

Mean cutoffs for average emotions can be seen in Table 4 for each group and visit. As the table shows, there did not seem to be any difference between visits or groups. The paired t-test on the collapsed emotions (happy, angry, sad combined) across visits found no significant differences in cutoff scores between visit 1 and visit 2. There were six participants in this data set.

Mean cutoffs for man versus woman faces can be seen in Table 4 for each group and visit. As the table shows, there did not seem to be any difference between visits or groups. The paired t-test found no significant difference in cutoff scores between visit 1 and visit 2. There were six participants in this data set.

Chi Square Analysis:

For our chi-square analysis, we included the average number of shifts by the TD and ASD groups. When looking at the raw number of participants who shifted up at least 3 times, stayed the same at least 3 times, and other, the table shows that there did not seem to be any difference between visits or groups for the raw participant data. This lack of pattern was confirmed by the chi-square statistic was 2.26 ($p = 0.32$). This means that we did not find a significant difference between the ASD and TD groups with respect to the number of trials shifting. However, the TD and ASD groups had different numbers of participants, and for that reason, we also did a chi square analysis on the percentages of participants who shifted according to each pattern. The table shows that when converted to percentages, there seems to be a difference in the shifts for the autistic participants versus the TD participants (with TD participants showing .33% shift up and .53% same and ASD participants showing .50% shift up and .25% same), confirmed by the chi-square statistic was 17.32, $p < .001$. This means that the groups differed in the distribution of participants who shifted in the different ways, between visit 1 and visit 2 (Figure 12). In particular, a higher proportion of the ASD group shifted their cutoff score upwards from visit 1 to visit 2, compared to the TD group.

Discussion

In this study, both TD and ASD groups shifted their cutoff values for both physical and emotion terms, from visit 1 to visit 2, once the 4 maximal pictures were added to the stimuli. This is a positive finding, and shows that these children were not significantly different from their TD peers at adolescence. Autistic and TD participants performed similarly to each other in this study, which is a change from earlier studies of younger participants. Some results even suggest that autistic individuals may be better at some forms of emotion and semantic recognition in our study; that is, the shift data shows that these autistic participants shifted more (upwards, following addition of 4 maximal stimuli) than the TD participants, which may suggest that these autistic individuals are better at emotion recognition, or at the very least better at the activity we studied. That is, taking the context into consideration more. Of course, this study was done with a very small sample size, and therefore the results may not be generalizable to the entire autistic population. However, this research does point to positive outcomes for autistic individuals in terms of emotion recognition as they age.

One interesting finding was that of the difference between male and female faces. We found that participants were more exclusive (meaning the mean cutoff values were higher on average) when it came to stimuli of male faces. This was an unexpected finding, and may be attributable to the greater sociocultural context in which we live. Men show emotion less frequently than women, on average, and this may account for the difference we are seeing in cutoff values for male versus female faces. Further research would be required to determine whether this is in fact the cause of the discrepancy.

There are a couple of reasons that could be behind the apparent improvement of autistic individuals as they age. For one, the age of the participants could affect their emotion and semantic recognition. There is potential for emotion and semantic recognition to improve as children's brains develop more fully. However, this improvement could also be caused by interventions in school and therapy. Many interventions are aimed at improving emotion and semantic recognition, in order to help autistic individuals function better in society at large (Baron-Cohen, 2009).

An asset of longitudinal studies is that data can be looked at over time, and we are able to see the improvement as participants age. As we can start to see, the previous de Villiers et al. study (which is a part of the same longitudinal study as the current study), where an almost identical method was used, found that children with ASD had significantly more difficulty sorting the emotion faces than their TD peers. However, in the current study, we see a very different picture. While this particular study did not compare the participant data to the de Villiers et al. (2019) participant data, this apparent change still illustrates the importance of longitudinal research because we can show how individuals change over time. Of course, detailed analysis would have to be done on both samples in order to determine significance. Moreover, the difference in methods, that the children saw the two stimulus sets during one visit when they were younger, but during two visits when they were older, cannot be ruled out as instrumental in the different results. Additionally, the visit 8 data used only the non-lenient coding, whereas the current study employed lenient coding, which could change the results. A further question that can be answered by continuing this longitudinal research is whether there is a point where autistic individuals will plateau, or whether they will continue to improve in their emotion and semantic recognition.

It is important to recognize that there could be other reasons for the improvement we see within the autistic individuals, other than just aging. For one, autistic individuals are taught how to sort pictures just like in our study as part of ABA (Applied Behavioral Analysis) and other types of therapy (7, 2022). This means that rather than truly understanding the meaning of the emotions on the faces, they may have just recognized how to memorize them. If this is the case, even as they improve in the lab, they may not be able to apply this improvement to the real world. In the real world, we would have to question whether they would resist calling a kindergartener tall because the prototype for 'tall' might be Lebron James. This is important to unpack as real world recognition is what these studies are trying to estimate. Determining whether this is the case would require a different experimental setup in order to control for this type of learning. Additionally, our study includes very few females, which is an issue concerning generalizability. This lack of female participation in autism research is widespread, and stems in part from the underdiagnosis of autistic females. Due to this, the findings from this study cannot be generalized to autistic women.

In conclusion, we have found that in these participants, the autistic participants did not differ significantly from the TD participants in terms of emotion recognition between visit 1 and visit 2. The autistic individuals did shift up significantly more than the TD participants on the shift analysis.

This research has the potential for importance within schools, and in developing better interventions for children with autism spectrum disorder. It will be important to study which interventions garner the most improvement in emotion and semantic recognition in order to best support autistic individuals.

Acknowledgements

We thank all the children, teens, and their families who participated in this research. This research was funded by the National Institute on Deafness and Other Communication Disorders (NIHDCD R01DC016665 ‘Early Predictors to School Age Language: Individual and Interactional Child and Parent Factors’).

A special thank you to Juandiego Carmona, Cynthia Boo, Kaya LeGrand, Amanda Mankovich, Lee Tecoulesco, and especially Dr. Letitia Naigles for all their wonderful support throughout the entire process of writing this paper.

References

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

Barner, David. Personal communication. November 19th, 2021.

Baron-Cohen, S., Golan, O., & Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions?. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1535), 3567–3574. https://doi.org/10.1098/rstb.2009.0191

Britton, M., Ipesa-Balogun, H., Jackson, E., Kotfila, J., Lu, K., Prabhakar, V., Torres, T. and de Villiers, P., 2017. *Acquisition of the Distinction Between Absolute and Relative Gradable Adjectives: Four-Year-Olds Know Emotion Adjectives are Relative*.

Diagnostic Criteria. (2022, April 6). *Centers for Disease Control.* Retrieved from https://www.cdc.gov/ncbddd/autism/hcp-dsm.html?msclkid=ecb0a2fec4b211ecaa44b52b13efc88 5

de Villiers, P., Naigles, L. and Tecoulesco, L.,(2019). *Children with Autism Distinguish between Absolute and Relative Gradable Adjectives: But not for Emotion Adjectives*.

Ellawadi, A. B., Fein, D., & Naigles, L. R. (2016). Category structure and processing in 6-year-old children with autism. *Autism Research, 10*(2), 327-336. doi:10.1002/aur.1652

Elliott, C. D. (2007). *Differential Ability Scales* (2nd ed.). San Antonio, TX: Harcourt Assessment

Foushee, R. and Srinivasan, M., 2017. Could both be right? Children’s and adults’ sensitivity to subjectivity in language.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr, Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, 30(3), 205–223.

Naigles, L., & Chin, I. (2015). Language in children with autism spectrum disorders. In *CUP Handbook* (pp. 637-658).

Naigles, L. R., & Fein, D. (n.d.). Looking through their eyes: Tracking early language comprehension in ASD. *Innovative Investigations of Language in Autism Spectrum Disorder.,* 49-69. doi:10.1037/15964-004

Naigles, L. R., & Tek, S. (2017). 'Form is easy, meaning is hard' revisited: (re) characterizing the strengths and weaknesses of language in children with autism spectrum disorder. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*(4). doi:10.1002/wcs.1438

Syrett, et al. (2006), *The Semantic Typology of Gradable Adjectives: Experimental Evidence from Adult and Child Language.*

Wiig, E. H., Semel, E., Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5).* Bloomington, MN: NCS Pearson.

Wit, T. C., Falck-Ytter, T., & Hofsten, C. V. (2008). Young children with Autism Spectrum Disorder look differently at positive versus negative emotional faces. *Research in Autism Spectrum Disorders, 2*(4), 651-659.

7 Autism Behavior and Communication Strategies. (2022). National University. Retrieved from https://www.nu.edu/resources/7-autism-behavior-and-communication-strategies/

Tables

Table 1

Means (SD) for demographic data by group

	Mean age in years (SD)	Number of participants (# Female)	Mean CELF WC Standard (SD)	Mean CELF FS Standard (SD)	Mean ADOS (SD)	Mean DAS (SD)
TD	14.87 (3.07)	15 (5)	11.86 (2.85)	11.86 (4.84)	1.93 (3.06)	109.60 (14.66)
ASD	15.33 (3.17)	12 (1)	6.25 (2.86)	6.08 (3.99)	13.67 (7.83)	84.75 (20.63)
Sig.	—	—	p=<0.001	p=<0.001	p=<0.001	p=<0.001

WC stands for Word Class subscore of the CELF test
FS stands for Formulated Sentences subscore of the CELF test

Table 2

Stimuli

Stimuli	**Dimension**	**Type of adjective**
Spotted dogs	Control, Physical	Absolute
Striped umbrellas	Control, Physical	Absolute
Long pencils	Control, Physical	Relative
Angry Woman	Emotion	Relative
Angry Man	Emotion	Relative
Sad Woman	Emotion	Relative
Sad Man	Emotion	Relative
Happy Woman	Emotion	Relative
Happy Man	Emotion	Relative

Table 3

Lenient data mean cutoffs and significance

Stimuli	**Mean Cutoff ASD V1 (SD)**	**Mean cutoff ASD V2 (SD)**	**Mean Cutoff TD V1 (SD)**	**Mean Cutoff TD V2 (SD)**	**Main effect of visit (p)**	**Main effect of group (p)**	**Group-visit interaction (p)**
Spotted Dogs	5 (0)	5 (0)	5 (0)	5 (0)	—	—	—
Striped Umbrellas	4.75 (0.9)	4.83 (0.98)	5 (0)	5 (0)	0.272	0.371	0.272
Long Pencils	6.04 (0.5)	6.31 (0.47)	5.88 (0.58)	6.26 (0.54)	0.001*	0.580	0.584
Overall Angry	4.95 (.94)	5.88 (1.09)	4.93 (1.31)	5.33 (1.75)	<0.001*	0.568	0.144
Angry Woman	4.5 (1.18)	5.92 (1.54)	5.13 (1.44)	5.53 (1.82)	—	—	—
Angry Man	5.27 (1.25)	5.83 (1.27)	4.73 (1.49)	5.13 (1.73)	—	—	—
Overall Sad	4.46 (1.27)	4.88 (1.40)	4.13 (1.09)	4.53 (1.52)	0.137	0.457	0.975
Sad Woman	4.42 (1.97)	4.75 (1.86)	3.6 (1.16)	4.13 (1.79)	—	—	—
Sad Man	4.5 (1.21)	4.91 (1.38)	4.67 (1.22)	4.93 (1.37)	—	—	—
Overall Happy	5.17 (1.27)	5.67 (1.51)	5.77 (.82)	6.03 (1.61)	0.075	0.316	0.577
Happy Woman	4.92 (1.25)	5.25 (1.60)	5.4 (1.34)	5.93 (1.66)	—	—	—
Happy Man	5.42 (1.80)	6.08 (2.19)	6.13 (0.77)	6.13 (1.77)	—	—	—

Table 4

Non-lenient data mean cutoffs and significance

Stimuli	**Number of total participants (out of 27)**	**Mean Cutoff ASD V1 (SD)**	**Mean Cutoff ASD V2 (SD)**	**Mean Cutoff TD V1 (SD)**	**Mean Cutoff TD V2 (SD)**	**Number of participants in analysis (out of 27)**	**Main effect of visit (p)**	**Main effect of group (p)**	**Visit-group interaction (p)**
Spotted Dogs	26	5 (0)	5 (0)	5 (0)	5 (0)	27	—	—	—
Striped Umbrellas	24	4.77 (0.90)	4.73 (0.95)	5 (0)	5 (0)	27	0.272	0.371	0.272
Long Pencils	27	6.05 (0.52)	6.31 (0.47)	5.88 (0.56)	6.26 (0.54)	27	0.001*	0.580	0.584
Overall Angry	12	5.00 (1.22)	5.88 (0.95)	4.63 (1.16)	4.50 (1.22)	12	0.022	0.239	0.005
Angry Woman	15	4.57 (0.79)	5.89 (1.16)	5.27 (1.42)	4.6 (1.35)	—	—	—	—
Angry Man	22	5.25 (1.25)	5.80 (1.32)	4.71 (1.44)	5.13 (1.73)	__	—	—	—
Overall Sad	9	3.33 (0.29)	3.67 (0.29)	3.42 (0.20)	3.58 (0.38)	9	0.170	1.000	0.626
Sad Woman	12	3.83 (1.60)	4.25 (1.57)	3.13 (0.35)	3.5 (1.27)	—	—	—	—
Sad Man	12	4 (0.5)	4 (0)	3.91 (0.30)	4.27 (0.47)	__	—	—	—
Overall Happy	18	4.75 (1)	5.33 (1.58)	5.85 (0.85)	6.00 (1.30)	18	0.047	0.065	0.814
Happy Woman	20	4.6 (1.06)	5.25 (1.60)	5.5 (1.39)	5.85 (1.66)	—	—	—	—
Happy Man	19	5.3 (1.85)	6.1 (2.45)	6.08 (0.83)	6 (1.38)	__	—	—	—

Table 5

Chi-square analysis: raw participant data

	3+UP	**3+same**	**other**
TD	5	8	2
ASD	6	3	3

The chi-square statistic is 2.2582. The p-value is 0.323327. The result is *not* significant at $p<0.05$.

Table 5A

Chi-square analysis: Percentages

	3+UP	**3+same**	**other**
TD	0.33%	0.53%	0.13%
ASD	0.50%	0.25%	0.25%

The chi-square statistic is 17.3181. The p-value is 0.000174. The result is significant at $p<0.05$.

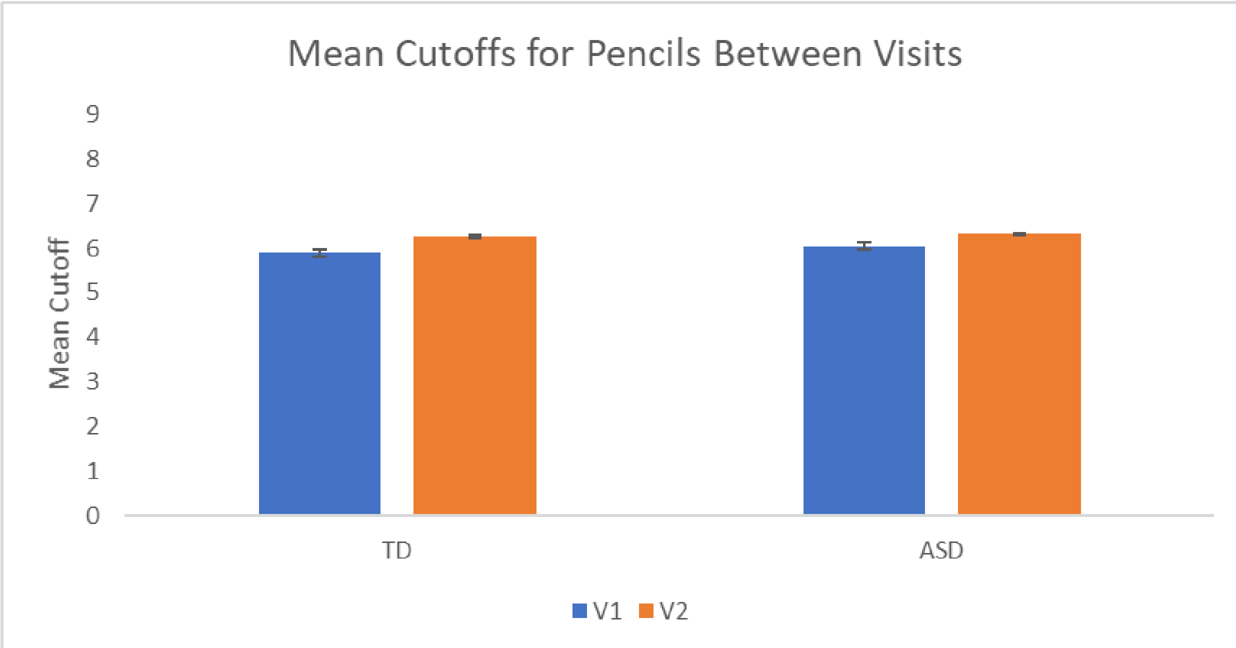

Figure 6: Mean cutoffs for pencils between visits

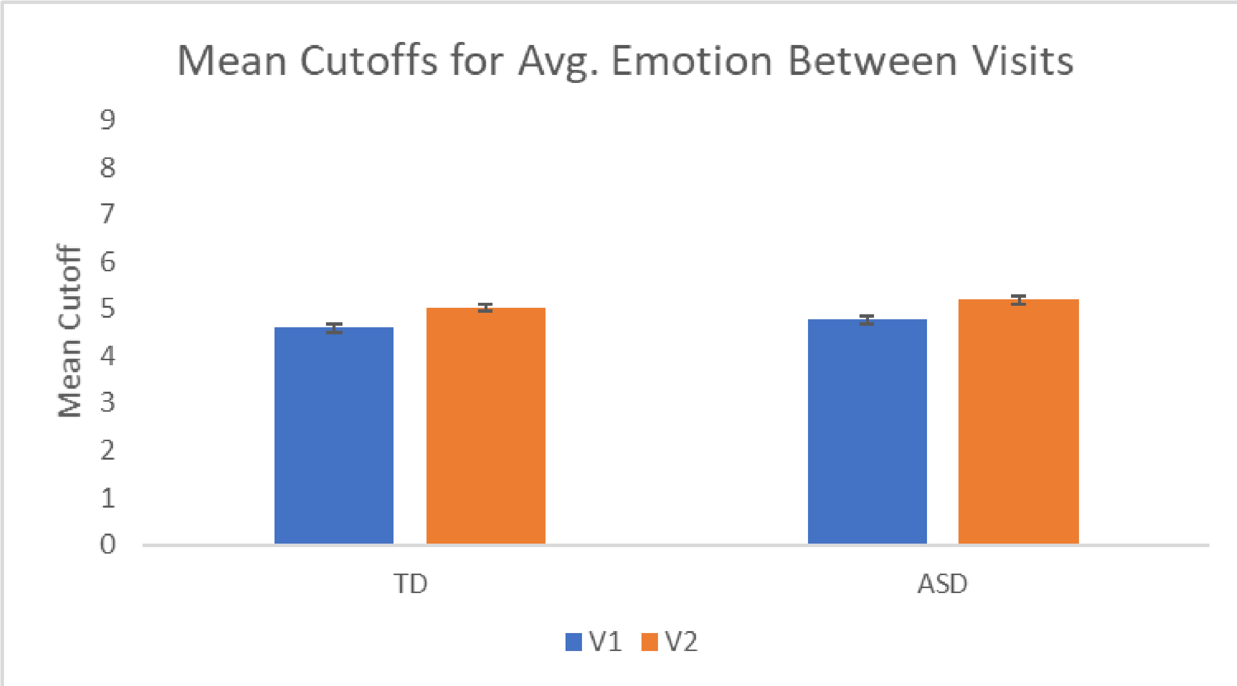

Figure 7: Mean cutoffs for average emotions between visits

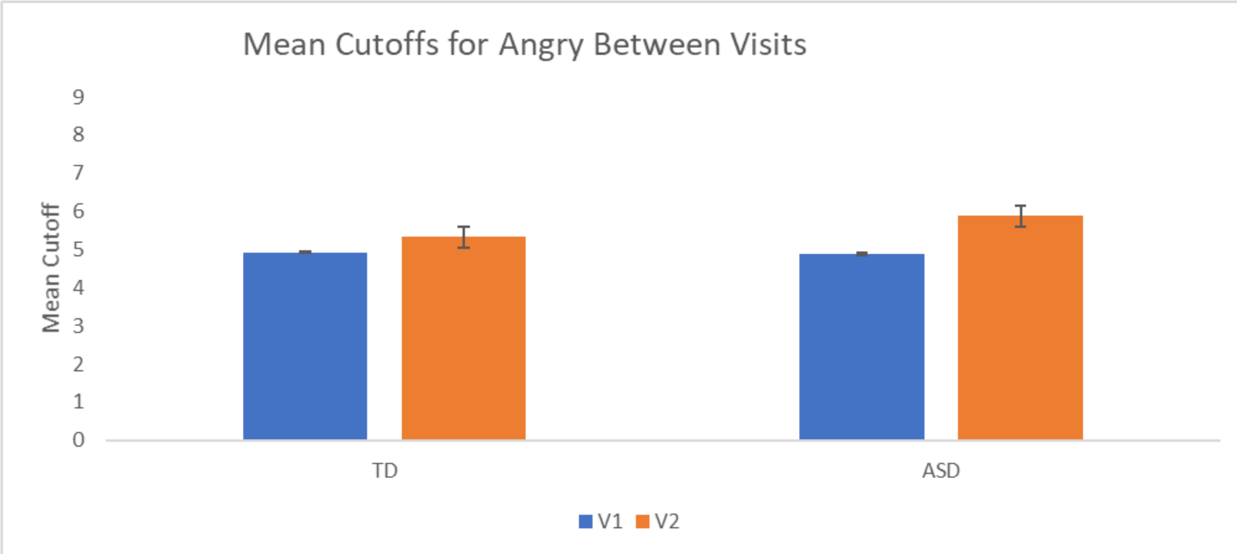

Figure 8: Mean cutoffs for angry faces between visits

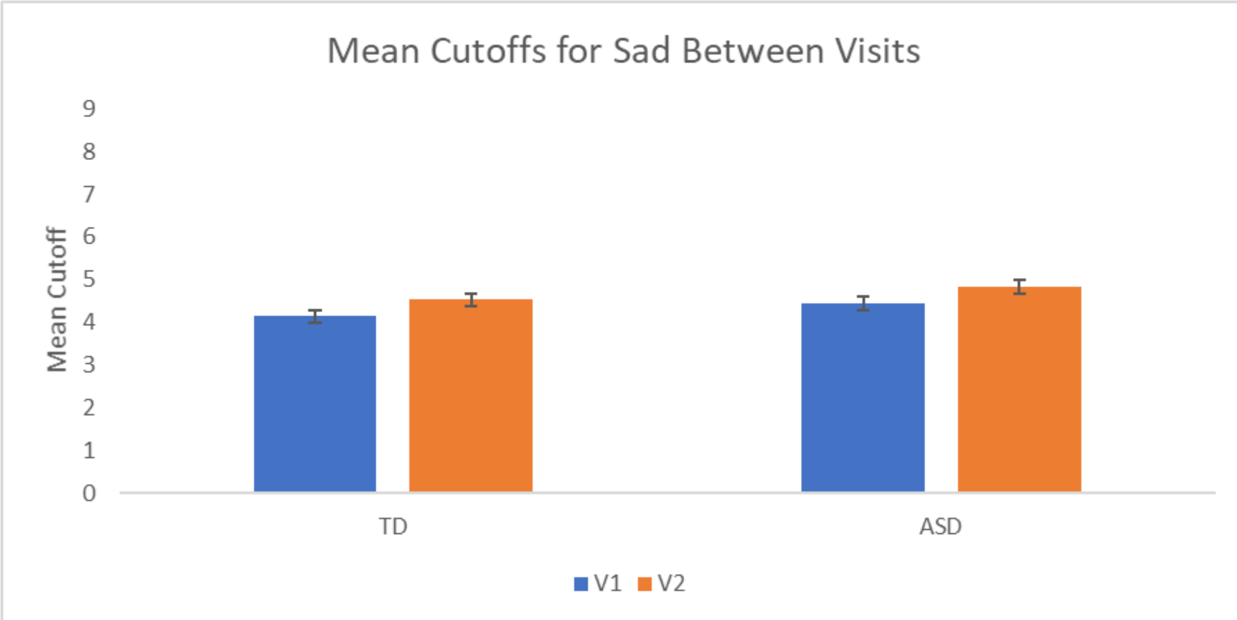

Figure 9: Mean cutoffs for sad faces between visits 1 and 2

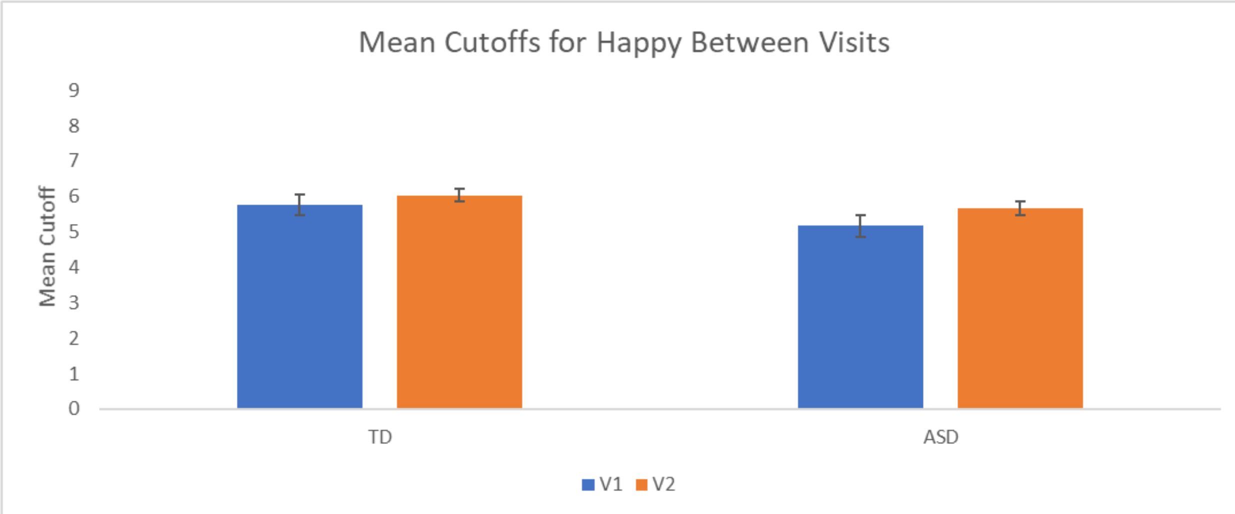

Figure 10: Mean cutoffs for happy faces between visits 1 and 2

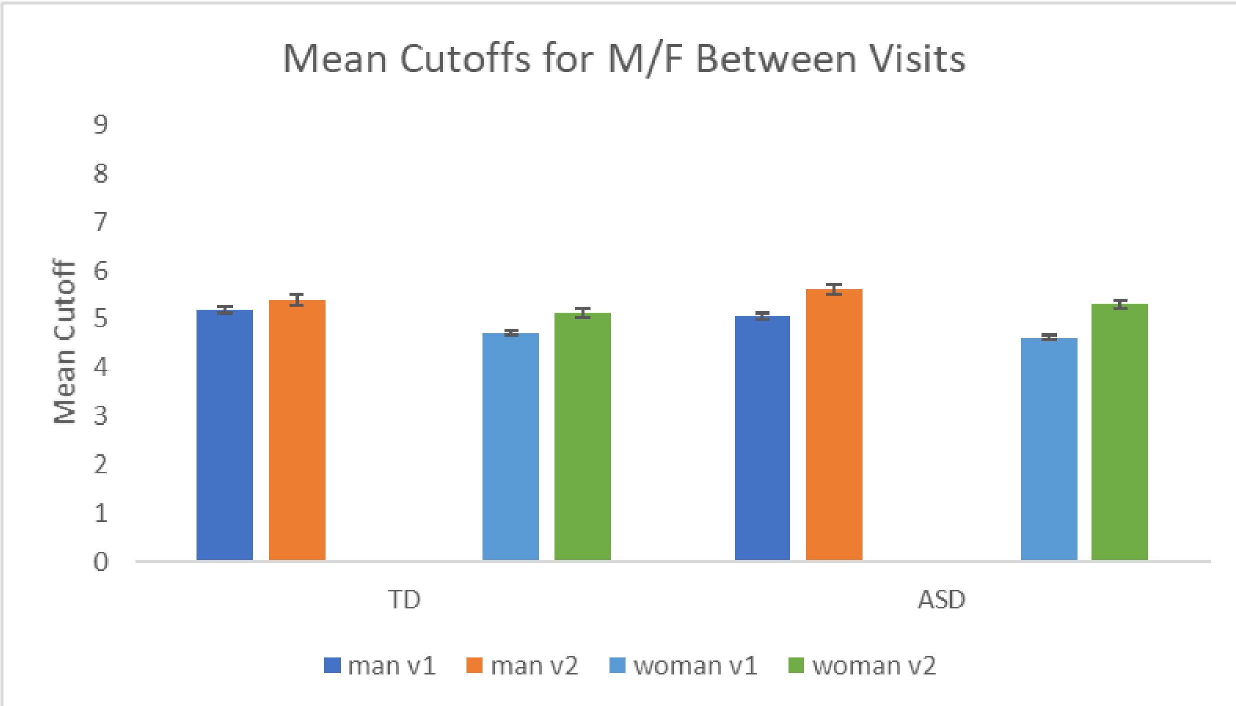

Figure 11: Mean cutoffs for male/female faces between visits 1 and 2

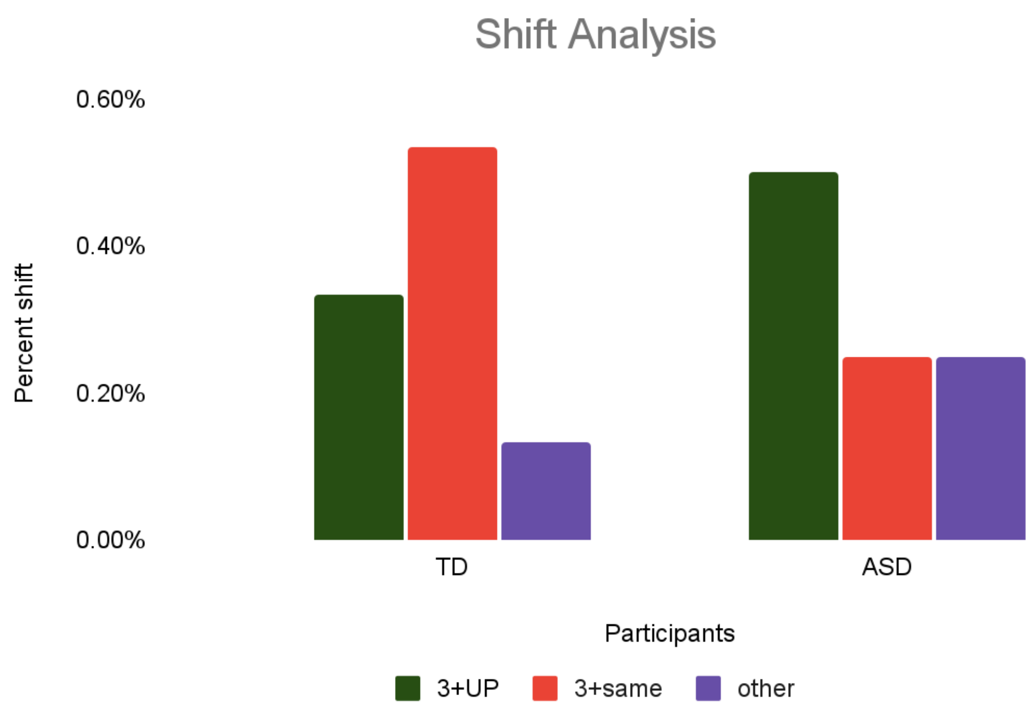

Figure 12: Shift analysis for TD and ASD groups