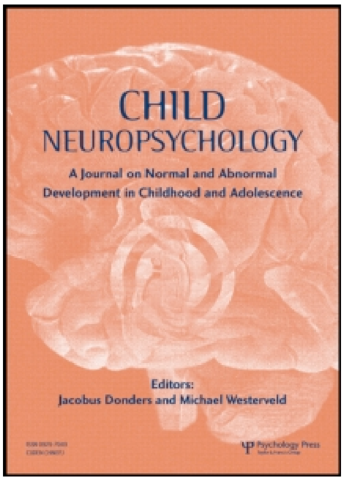

Child Neuropsychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713657840

The Effect of Music on Social Attribution in Adolescents with Autism Spectrum Disorders

Anjali K. Bhatara [a]; Eve-Marie Quintin [b]; Pamela Heaton [c]; Eric Fombonne [ad]; Daniel J. Levitin [a]
[a] McGill University, Montreal, Canada [b] Université du Québec à Montréal, Canada [c] Goldsmiths College, University of London, United Kingdom [d] Department of Psychiatry, Montreal Children's Hospital, Montreal, Canada

First Published on: 13 January 2009

To cite this Article Bhatara, Anjali K., Quintin, Eve-Marie, Heaton, Pamela, Fombonne, Eric and Levitin, Daniel J.(2009)'The Effect of Music on Social Attribution in Adolescents with Autism Spectrum Disorders',Child Neuropsychology,15:4,375 — 396

To link to this Article: DOI: 10.1080/09297040802603653

URL: http://dx.doi.org/10.1080/09297040802603653

Child Neuropsychology, 15: 375–396, 2009
http://www.psypress.com/childneuropsych
ISSN: 0929-7049 print / 1744-4136 online
DOI: 10.1080/09297040802603653

Psychology Press
Taylor & Francis Group

THE EFFECT OF MUSIC ON SOCIAL ATTRIBUTION IN ADOLESCENTS WITH AUTISM SPECTRUM DISORDERS

Anjali K. Bhatara,[1] Eve-Marie Quintin,[2] Pamela Heaton,[3] Eric Fombonne,[1,4] and Daniel J. Levitin[1]

[1]McGill University, Montreal, Canada, [2]Université du Québec à Montréal, Canada, [3]Goldsmiths College, University of London, United Kingdom, and [4]Department of Psychiatry, Montreal Children's Hospital, Montreal, Canada

High-functioning adolescents with ASD and matched controls were presented with animations that depicted varying levels of social interaction and were either accompanied by music or silent. Participants described the events of the animation, and we scored responses for intentionality, appropriateness, and length of description. Adolescents with ASD were less likely to make social attributions, especially for those animations with the most complex social interactions. When stimuli were accompanied by music, both groups were equally impaired in appropriateness and intentionality. We conclude that adolescents with ASD perceive and integrate musical soundtracks with visual displays equivalent to typically developing individuals.

Keywords: ***Autism; Music; Social attribution; Pervasive developmental disorders; Asperger syndrome.***

Among typical adults, it is well established that the presence of music (a "musical soundtrack") can influence the emotional impact of a visual scene or video game (Cohen, 2001; Lipscomb & Zehnder, 2004) and even alter perception and later memory of events in a visual scene (Boltz, 2001, 2004; Boltz, Schulkind, & Kantra, 1991). There is less evidence of this effect in children, but children's interpretations of stories do vary according to the emotional valence of background music presented with the story (Ziv & Goshen, 2006). To date, no research has addressed the question of how musical soundtracks influence perception and cognition in individuals with Autism Spectrum Disorders (ASD). These disorders are characterized by abnormalities in social cognition. The main goal of the present study was therefore to determine whether the presence of a musical soundtrack affects social cognition differentially in children with ASD and typically developing controls. If true, it would indicate that children with ASD are either

The research reported herein was submitted in partial fulfillment of the requirements for the PhD in Psychology at McGill University by the first author. Dr. Anjali Bhatara is now at the Department of Surgery, Division of Head & Neck, University of California, Los Angeles. The research was funded by grants to DJL from NSERC, NAAR, and NSF. We are grateful to Bennett Smith and Karle-Philip Zamor for technical assistance in programming the experiment and preparing the stimuli, to Bianca Levy for help with testing participants and for coding responses, to Vanessa Park-Thompson for coding, and to Athena Vouloumanos for valuable comments.

Address correspondence to Daniel J. Levitin, PhD, Department of Psychology, McGill University, 1205 Avenue Penfield, Montreal, QC H3A 1B1, Canada. E-mail: daniel.levitin@mcgill.ca

processing the music differently or integrating the music with the animation in different ways from the controls. If, however, the music affects the two groups in the same way, then these findings could be taken as additional evidence for music as a spared domain of cognitive processing in ASD.

Autism is a heterogeneous neurodevelopmental disorder defined by behavioral characteristics (*Diagnostic and Statistical Manual of Mental Disorders*, fourth edition, text revision [*DSM-IV-TR*]; American Psychiatric Association, 2000). There are three definitional criteria: deficits in social behavior, impaired communication abilities, and a restricted range of interests. In order to be diagnosed with classic autism, the patient must fulfill all three criteria outlined above. Often, he or she will show language delay. If the patient shows deficits in social behavior, has a restricted range of interest and has a high enough IQ and no language delay, he or she may be diagnosed with Asperger syndrome. If only some of the three criteria are fulfilled, and the patient does not fit the criteria for Asperger syndrome, a clinician may diagnose him or her with pervasive developmental disorder – not otherwise specified (PDD-NOS). All of these disorders are part of the autism spectrum, with large variation in the severity of symptoms (Wing, 1997; Wing & Gould, 1979). Current estimates of the prevalence of ASD in the population suggest that between 6 and 7 out of 1,000 individuals are affected with an ASD, with the highest proportion of these being classified as PDD-NOS (Fombonne, 2005; Fombonne, Zakarian, Bennett, Meng, & McLean-Heywood, 2006). Because of the heterogeneity of ASD and the relative neglect of individuals with PDD-NOS in research studies, our study was not focused solely on investigating classic autism but was broadened to include children and adolescents spanning the entire spectrum of autism, Asperger syndrome, and PDD-NOS.

While individuals with classic autism, Asperger syndrome, and PDD-NOS meet slightly different criteria according to *DSM-IV*, they share several common elements. Difficulties in three areas are broadly associated with the autism spectrum: the ability to predict or interpret others' thoughts, moods, and actions (Leslie, 1987); social cognition (defined as the ability to select, to interpret, and to use social information to make judgments and decisions; Taylor, 1997); and the ability to interpret affective cues of differing complexity (Capps, Yirmiya, & Sigman, 1992; Hobson, Ouston, & Lee, 1989).

Among typically developing children and adults, there is generally strong agreement as to which of the basic emotions is portrayed by a given piece of music (Cross, in press), both within and across cultures (Balkwill, Thompson, & Matsunaga, 2004). Note, however, that data from a study carried out with children and musically naïve adults showed considerably less consensus about the meaning of complex emotions in music such as *triumph* and *contemplation* (Heaton, Allen, Williams, Cummins, & Happé, 2008). Nonetheless, this is still a useful approach for examining social cognition of children with ASD. In this method, researchers ask children with ASD to identify pictorially represented emotions and mental states by pairing them with extracts of music depicting these mental states and emotions. In the first of these studies, Heaton et al. (1999) showed that children with autism were unimpaired in recognizing two basic emotions (happiness and sadness) when they were associated with music in major and minor modes. In a second study, Heaton et al. (2008) extended this paradigm to include measures of more complex emotions and mental states (fear, anger, tenderness, triumph, and contemplation). They tested a large sample of typically developing 4- to 10-year-old children and found that chronological age accounted for most of the variance in performance. When intellectually high- and low-functioning children and adolescents with autism, and adolescents and young adults with Down syndrome completed the same task, there was no significant effect of diagnosis once verbal mental age was factored out of the analysis. This suggests that children with autism who achieved sufficient verbal mental age scores showed no deficits in understanding and categorizing the stimuli. An important question arising from the findings of emotion recognition in music is whether such abilities may be used to increase the poor levels of social understanding characteristic of this disorder.

An influential theoretical account of the typical deficits in social cognition that characterizes ASD is that individuals with ASD are impaired in "Theory of Mind" (ToM; Baron-Cohen, Leslie, & Frith, 1985), the ability of individuals to connect others' observable behaviors with their internal mental states (Bauminger & Kasari, 1999) in order to predict their behavior. The ability to understand others' beliefs and intentions is often tested using False Belief tasks (for a review and meta-analysis of these, see Wellman, Cross, & Watson, 2001). Research has shown that typically developing children pass these tests by age 4 or 5 (Sabbagh, Moses, & Shiverick, 2006), but children with autism frequently do not pass these tests (see Baron-Cohen, 2001, for a review). There is evidence that those who do pass use strategies that differ from those used by typically developing (TD) children (Fisher, Happé, & Dunn, 2005). However, the extent to which ToM tasks predict or reflect functioning outside of the experimental setting has become a matter of considerable debate. Klin, for example (2000), presented data from a sample of intellectually unimpaired individuals with autism or Asperger syndrome. Most of these individuals possessed sufficient language skills to be able to pass False Belief tasks in the lab but were highly impaired in interpersonal relationships as measured by the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Dicchetti, 1984). This finding is consistent with numerous other studies showing that success on False Belief tasks does not preclude characteristic deficits in social interactions in autism (e.g., Bauminger & Kasari, 1999; Ozonoff & Miller, 1995). Klin points out that these laboratory tasks do not measure spontaneous seeking of social information in the environment, the ability to distinguish important from irrelevant social information, or the ability to integrate environmental information into a social context. This suggests that the construct tested in most ToM tasks is only one of many social cognitive mechanisms, any of which can impair functioning in the real world. For example, one mechanism of interest is the ability to attribute social intention to animated figures.

Klin (2000) and Bowler and Thommen (2000) adapted Heider and Simmel's (1944) animation of geometric figures to examine the ability of children and young adults with ASD to understand nonverbal social communication and to attribute social intentions and actions to the moving figures. Klin found that while the control group reacted to the animations much like Heider and Simmel's (1944) adult participants, the young adults in the ASD group failed to identify many of the social elements of the story and made fewer relevant social attributions. Bowler and Thommen found that children with autism described fewer interactions between the figures when compared with chronological-age- and IQ-matched typical control groups, and they also involved themselves more in their descriptions of the animation than the children without autism (e.g., "*I'm* going to catch that grey . . .").

Abell, Happé, and Frith (2000) also used this paradigm with high-functioning children with ASD, but rather than using Heider and Simmel's (1944) animation, which mainly portrayed goal-directed action rather than mental-state attribution, Abell et al. developed several shorter animated cartoons with only two figures interacting. They created three categories of interaction: Random (Rnd), Goal-directed (GD), and Theory of Mind (ToM). In the random animations, two triangles moved in random fashion and did not affect each other's movement. In the GD animations, the two triangles appeared to respond to each other's behavior thereby demonstrating intentionality. However, their putative intentions were to perform physical actions (e.g., chasing, leading, fighting) and so did not involve "mind-reading." In the ToM animations, the two triangles interacted in more complex ways, with one triangle demonstrating an intention to influence the other triangle's mental state (e.g., coaxing, mocking, surprising).

After viewing each animation, the participants were asked to describe what they saw happening in the animation. These responses were rated on two dimensions: overall accuracy and type of description. Accuracy was rated by comparing the participant's description with the animation designer's intended meaning. There were three categories for types of description: whether they described simple actions, interaction without mental state terms (interaction), or interactions with mental-state terms (mentalizing). The authors found that there was a trend toward a group difference among the autism and control groups, with the autism group being less accurate overall. In addition, fewer participants in the autism group achieved perfect scores on at least two of the four ToM animations than the other groups. There were no differences between groups for Rnd and GD. It was also noted that the autism group used significantly fewer appropriate mentalizing descriptions for the ToM animations than the control group. The most salient finding from the study was the difference in levels of appropriateness of mentalizing descriptions between groups.

Castelli, Frith, Happé, and Frith (2002) used these same animations in a neuroimaging study of adults with and without autism or Asperger syndrome to examine the neural bases of these types of social attribution. They found similar behavioral results to Abell et al. (2000), with the autism group giving less accurate descriptions and fewer mental-state descriptions of the ToM animations, but not varying from the control group on GD or Rnd animations. Thus, deficits observed in this study were specific to the ToM animations.

Findings from experiments testing social cognition using music (Heaton et al., 1999, 2008) suggest that music provides emotional and mental-state cues that children with autism can understand. In addition, experiments adding music to films (Boltz, 2001, 2004) or Heider and Simmel's (1944) animation (used by Marshall & Cohen, 1988) have shown that typical adults' perception of events in the sequence changes in a specific way as a result of the music. In order to investigate these phenomena in individuals with ASD, we added musical soundtracks to animations developed and used by Abell et al. (2000) and Castelli et al. (2002) and presented both these and silent versions to adolescents with and without ASD and a control group of age- and IQ-matched typically developing adolescents. We compared the results from the ASD group to those from the control group, and we also examined results from diagnostic subgroups within the ASD group (PDD-NOS vs. Asperger syndrome).

In the present experiment, the presence of music with the animations is predicted to modify the typically developing (TD) control group's interpretation of the events. The rationale for this prediction draws on studies by Boltz (2001, 2004) and Cohen (2001) showing modification of interpretation of visual scenes in response to music. Further, because of music's emotional force, we predict that the descriptions of those animations with the greatest social attribution component will be most influenced by the presence of music, that is, the ToM cartoons followed by the GD cartoons.

With respect to diagnosis, our primary hypothesis (H_1) is that deficits in social attribution are robust and cannot be influenced by the addition of music. The alternative hypothesis (H_2) is that while individuals with ASD will be impaired in social attribution relative to matched controls, such deficits may nevertheless show a shift in responses when music is added to the animations. This is because social communicative deficits fall on a continuum in ASD, and empirical data suggest some degree of sparing of perception of emotions and mental states in music (Heaton et al., 1999, 2008). If our findings show that ASD participants' descriptions are influenced by the addition of musical soundtracks differently from those of controls (as measured by appropriateness or number of intentionality words), then this would be evidence that individuals with ASD are perceiving the music or the combination of the music and cartoon in a different way from controls.

METHOD

Participants

There were two experimental groups in this study. Initially, 33 children with Autism Spectrum Disorders (ASD) were recruited: 25 from a specialized autism clinic at the Montreal Children's Hospital and 8 from a private school for children with physical and mental disabilities. These participants were aged between 10 and 19 years and had all been diagnosed according to *DSM-IV* criteria by specialized medical teams with expertise in diagnosing autism and other ASDs. Subgroup diagnosis (Asperger syndrome and PDD-NOS) was done according to *DSM-IV* criteria. Forty-six typically developing children between the ages of 8 and 18 were recruited by word of mouth and from four public schools in Montreal.

For the study's analyses, 26 participants in the TD group (14 girls and 12 boys) were matched to a group of 26 participants with ASD (3 with Autism, 13 with Asperger syndrome, and 10 with PDD-NOS; 6 girls and 20 boys) such that verbal IQ (VIQ), performance IQ (PIQ) and full scale IQ (FSIQ) scores from the Wechsler Abbreviated Scale of Intelligence differed by less than one standard deviation (see below for descriptions of screening instruments; see Table 1 for results). The *t*-tests on chronological age and PIQ showed no difference between groups, $t_{age}(50) = 0.15$, $p = .89$, and $t_{PIQ}(50) = 1.59$, $p = .12$. Independent sample *t*-tests showed no difference between groups on years of musical

Table 1 Descriptive Statistics for Participants with ASD ($N = 26$) and Typically Developing (TD) ($N = 26$) Participants

	Age (yr:mo)	FSIQ	VIQ	PIQ
ASD (6 girls, 20 boys)				
Mean	13:7	97	94	101
SD	1:11	15	19	13
Range	10:10–19:4	79–133	62–132	81–129
TD (14 girls, 12 boys)				
Mean	13:6	108	107	107
SD	2:2	12	13	15
Range	9:11–17:9	79–132	81–133	75–137
t(50)	0.15	2.76**	2.91**	1.59

FSIQ: Full scale IQ, VIQ: verbal IQ, PIQ: performance IQ.
$^{**}p < .01$.

training, $t_{yrs}(48) = 0.86$, $p = .40$, number of instruments played, $t_{inst}(50) = 1.55$, $p = .13$, or scaled scores for digit span or letter-number sequencing, $t_{DS}(50) = 0.32$, $p = .75$, and $t_{LN}(50) = 1.38$, $p = .17$.

Background and Screening Measures

All participants completed the Wechsler Abbreviated Scale of Intelligence (WASI). As the experiment described below has considerable auditory-temporal demands, we also evaluated the participants on their auditory working memory with the Digit Span and Letter-Number Sequencing subtests of the Wechsler Intelligence Scale for Children-IV (WISC-IV). Musical experience was measured by participants' responses on a revised version of the Queens Questionnaire for Musical Background (Cuddy, Balkwill, Peretz, & Holden, 2005) and the parents' responses on the Salk and McGill Music Inventory (SAMMI; Levitin, Cole, Chiles, Lai, Lincoln, & Bellugi, 2004). The parents of the participants also completed two questionnaires about social functioning: the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) and the Social Responsiveness Scale (SRS; Constantino et al., 2003). These latter two instruments provided us with a way to screen our control participants for signs of autism (see Table 2). No participants in the TD group met the cutoff for ASD on the SCQ.

The parents of three of the participants in the ASD group reported on the SAMMI questionnaire that their child had hearing-related problems in early childhood, but that these problems were no longer present at the time of testing. One additional child reported impairment in one ear but was included in the study given that one ear with normal hearing is sufficient to complete the task. No hearing problems were reported for the TD group.

Stimuli

Stimuli were based on a paradigm initially developed by Heider and Simmel (1944). The animations used in the present study were those developed by Abell et al. (2000), as described above. These animations showed two triangles moving around the screen in one

Table 2 Descriptive Statistics for Participants with ASD ($N = 26$) and Typically Developing (TD) ($N = 26$) Participants

	Number of Instruments played	Years of musical experience (yr: mo)	SRS	SCQ
ASD				
(6 girls, 20 boys)				
Mean	1	2:2	75	18
SD	0.8	3:3	14	14
Range	0–3	0–12	43–90	1–29
TD				
(14 girls, 12 boys)				
Mean	1	2:11	49	4
SD	1.1	3:1	9	9
Range	0–4	0–11	36–73	0–13
$t(50)$	1.55	0.86	8.16**	9.4**

SRS: Social Responsiveness Scale, SCQ: Social Communication Questionnaire.
**$p < .01$.

of three different ways, which were described by Abell et al. as Random (Rnd; e.g., bouncing), Goal-directed (GD; e.g., fighting), or Theory of Mind sequences (ToM; e.g., mocking). We selected a subset of each of these three types of animations and added a unique musical soundtrack to each animation, thus creating two conditions, visual only (the original animations) and auditory-visual combined (with our added musical soundtracks).

Our original goal was to create one version with music that "matched" the animation well and one version that was not well matched. Thus, we initially created two musical versions of each of the 11 animations used in the Abell et al. (2000) study, resulting in a total of 22 musically accompanied animations. Sixteen adults were asked to rate each animation on a scale of 1 to 10, with 10 being "music matches well" and 1 being "music does not match." The pilot study revealed poor consensus and it appeared that, in many cases, either type of music could be perceived as well matched. Given that the geometric figures were relatively abstract, with no faces or bodies to communicate expression, the moods of the scenes themselves were more ambiguous than scenes with people acting or speaking. Therefore, there was much less of a visual scene for the music to match, so it is not surprising that the music showed no strong consensus.

Because of this lack of consensus, we included only one musical version for each animation—the one that obtained the higher matching rating. However, for two animations, one GD (*Leading*) and one ToM (*Mocking*), we included both musical versions because there was a significant difference in the ratings between the two different types of music, $t_{\text{Leading}}(15) = 2.57$, $p = .02$ and $t_{\text{Mocking}}(15) = 3.14$, $p = .007$. In order to keep the experiment to a manageable length of time, one animation from each category (Rnd, GD, and ToM) was eliminated. The criteria for this were redundancy (chasing being similar to leading, so we eliminated the former; all three random animations elicited similar responses) or lack of matching (neither musical selection we chose was rated as matching well with the coaxing animation).

This resulted in a total of 8 silent animations (2 Random, 3 Goal-Directed, and 3 Theory of Mind) and 10 music animations (2 Random, 4 Goal-Directed, and 4 Theory of Mind) with *Leading* and *Mocking* each paired with two different musical excerpts (for a list of these, see Appendix A. Stimuli are available online at http://www.psych.mcgill.ca/labs/levitin/triangles.htm). These were presented according to the procedure outlined below.

Procedure

Participants saw all versions of the animations in two blocks. Block one consisted of the 8 silent animations and block two consisted of the 10 animations that were accompanied by music. The order of presentation of blocks was counterbalanced, and the animations were presented in random order within the blocks.

Auditory stimuli were presented through loudspeakers (Acoustic Research, Model 570, Hauppauge, New York). The experiment was presented in Psiexp (Smith, 1995) on a Macintosh Powerbook. After each animation, participants rated how well they thought the music matched the animation on a slider scale on the computer screen. The left and right ends of the slider scale were labeled "Does not match at all" and "Matches very well," respectively. There were no numbers on the scale, but there were four tick marks along the edge of the scale so participants could better keep track of their own ratings. In addition, participants answered the question: "What do you think was happening in the cartoon?" Their responses were recorded using a Sony PCM-M1 portable DAT recorder and then transcribed. They were then coded by two independent raters using criteria modified from Castelli, Happé, Frith, and Frith (2000, received from F. Castelli, personal communication, June 15, 2006).

The independent variable was Diagnosis (ASD versus typically developing). The four dependent variables were Music Matching (how well the music matched the cartoon), Length of Description, Appropriateness of Description, and Presence of Intentionality. The Length score was determined by the number of clauses from 0 to 5, with 0 being no response and 5 being five or more clauses. The Appropriateness score reflects how accurately the participant's description of the event depicted in the animation matched the original intent of the designers of the animation. This was scored from 0 to 3, with 0 representing an "I don't know" response and 3 representing an accurate and complete description. The Intentionality score reflected the degree to which the participant described complex, intentional mental states and was scored on a scale of 0 to 5 with 0 representing no purposeful action and 5 representing intentional manipulation of the mental state of another (See Appendix B). As measured by two-way mixed-model intraclass correlation (ICC; Shrout & Fleiss, 1979), interrater agreement for all categories was high, with all three ICCs ≥ 0.89 ($p < .001$).

RESULTS

Music Matching

We performed a two-way repeated measures ANOVA on the Music Matching ratings with diagnosis as a between-subjects factor and animation type as a within-subject factor. We found that the diagnostic groups did not differ from each other in their ratings of how well the music matched the animation, $F(1, 50) = 0.18$, $p = 0.68$, indicating that children with ASD did not differ from typically developing controls in this measure. There were significant differences in music-matching ratings across the animation types, $F(2, 100) = 10.47$, $p < .001$. Bonferroni-adjusted pairwise comparisons among the animation types showed that all participants tended to rate the music excerpts as better matched with the Theory of Mind (ToM) animations than with the Goal-directed (GD), $t(51) = 2.90$, $p < .01$, and Random (Rnd), $t(51) = 4.08$, $p < .001$, types of animations. The difference between GD and Rnd was marginally significant, $t(51) = 1.89$, $p = .06$. There was no interaction between animation type and diagnosis. In sum, the two groups did not differ from each other in music matching, and all participants rated the music as better matched with the ToM animations than with the other two.

Length

We performed a three-way repeated measures ANOVA on length of description scores, with diagnosis as a between-subject factor and music/silent and animation type as a within-subject factor. The main effect of diagnosis was significant, $F(1, 50) = 4.02$, $p < .05$, because the ASD group provided longer descriptions for all animations. The main effect of animation type was also significant, $F(2, 100) = 78.32$, $p < .001$. Bonferroni-adjusted pairwise comparisons showed that all participants' descriptions of the ToM animations were longer than those of the GD, $t(51) = 5.48$, $p < .001$, and the descriptions of the GD animations were longer than the descriptions of the Rnd, $t(51) = 5.75$, $p < .001$. There was also a trend toward an interaction between animation type and diagnosis, $F(2,100) = 2.91$, $p = .06$, shown in Figure 1. While the control group showed a strong tendency to provide longer descriptions of the ToM animations than of the other two, the ASD group tended to provide

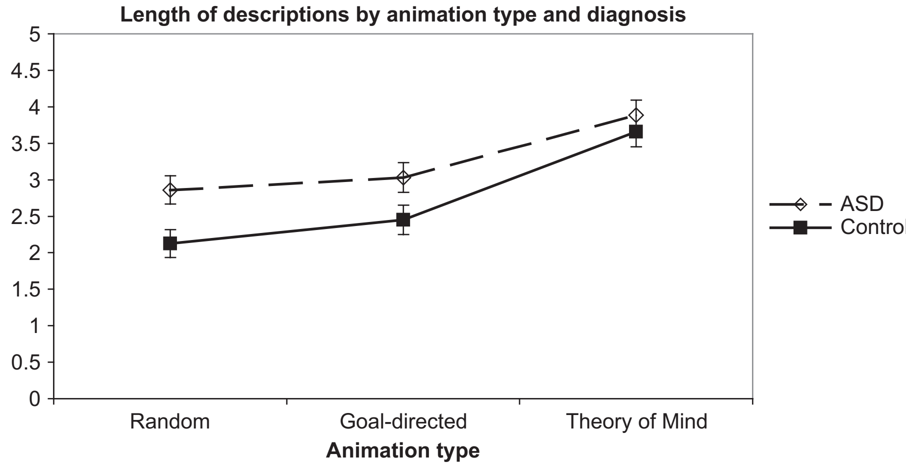

Figure 1 Length of descriptions (mean number of clauses) of the ASD and TD groups for each animation type.

descriptions of similar length across animation types. Bonferroni-adjusted pairwise comparisons showed that the ASD group's descriptions were significantly longer than the control group's for the Rnd animations, $t(50) = 3.33$, $p = .002$, and for the GD animations, $t(50) = 4.92$, $p < .001$, but did not differ significantly for the ToM animations, $t(50) = 0.75$, $p = .46$. Tukey's HSD post hoc tests showed that there was a significant difference between the length of descriptions of the GD and ToM animations for the ASD group ($p < .01$), but the descriptions of the Rnd and GD did not differ significantly in length. However, all three of these animation types differed in length for the control group ($p < .01$).

There was also a significant main effect for the music/silent condition, $F(1, 100) = 11.82$, $p = .001$, with all participants producing longer responses to the silent animations than to the animations with music.

In summary, all participants gave the longest descriptions for the ToM, next longest for the GD, and the shortest descriptions for the Rnd animations; though the difference between the GD and Rnd was not significant for the ASD group. The ASD group gave longer descriptions than the control group for the Rnd and GD animations, but the groups did not differ on the ToM animations. Also, all participants gave longer descriptions for the silent animations than for those with music.

Appropriateness

We performed a three-way repeated measures ANOVA on the Appropriateness scores with diagnosis as a between-subject factor and music/silent and animation type as within-subject factors. The main effect of diagnosis was not significant, $F(1, 50) = 2.3$, $p = .14$. The main effect of animation type was significant, $F(1, 50) = 15.86$, $p < .001$. Bonferroni-adjusted *t*-tests (with an adjusted of .01) showed that the ToM animations were described less appropriately than the GD and the Rnd animations, $t(51) = 4.35$, $p < .001$ and $t(51) = 5.21$, $p < .001$, respectively. Also, the GD animations showed a trend toward being described less appropriately than the Rnd animations, $t(51) = 2.15$, $p = .04$. In addition, the main effect of music/silent was significant, $F(1, 50) = 14.94$, $p < .001$. Animations with music were described less appropriately than silent animations, and each of the participants' appropriateness was affected by the music on at least one of the animations. There were no significant interactions among the three factors. In summary, the analysis of the appropriateness measure showed that both the presence of music and more complex animation types caused all participants' descriptions to become less appropriate.

Intentionality

We performed a three-way repeated measures ANOVA on Intentionality with diagnosis as a between-subject factor and music/silent and animation type as within-subject factors. The main effects of diagnosis and music/silent were not significant, but the main effect of animation type was significant, $F(2, 100) = 136.6$, $p < .001$. Bonferroni-adjusted comparisons (with an adjusted of .01) showed that the ToM animations' descriptions included more intentionality words than those of the GD animations, $t(51) = 9.58$; $p < .001$, and the GD animations' descriptions included more intentionality words than those of the Rnd animations, $t(51) = 9.76$, $p < .001$. There was a significant interaction between animation type and diagnosis, $F(2, 100) = 4.65$, $p = .01$, which indicates that for the ToM animations, as expected from our a priori hypothesis, the ASD group used fewer intentionality words than the control group, $t(50) = 2.47$, $p = .02$ (see Figure 2).

In addition, there was a significant two-way interaction between animation type and music/silent, $F(2, 100) = 5.55$, $p < .01$. This interaction shows that the participants' descriptions for the ToM animations contained words with less intentionality when there was music present than when the animations were silent (see Figure 3).

To further explore the relationship between ASD symptom severity and our dependent measures, we performed Pearson correlations between the SRS T-scores and the dependent variables. When both groups were combined, we found a significant negative correlation between SRS scores and intentionality scores for the ToM animations, $r = .34$, $p = .01$. However, when the diagnostic groups were analyzed separately, this correlation was no longer significant.

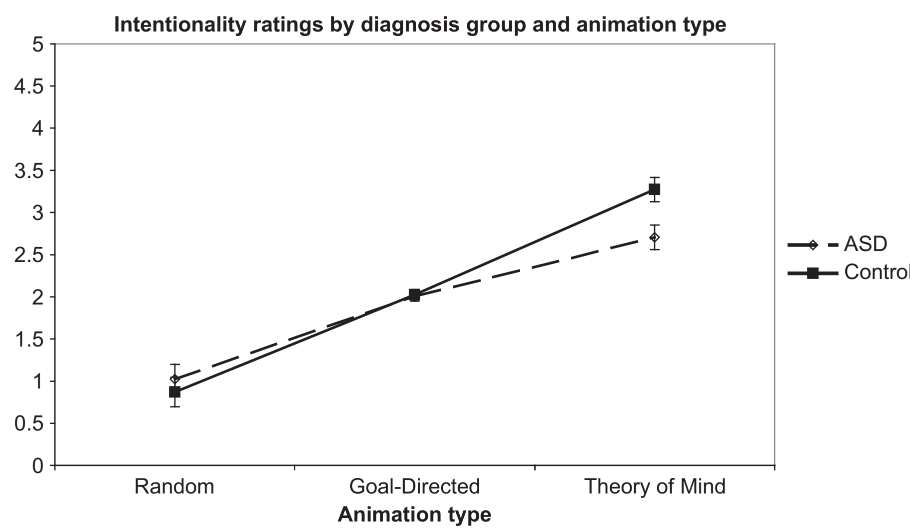

Figure 2 Intentionality ratings of responses of the ASD and TD groups for each animation type.

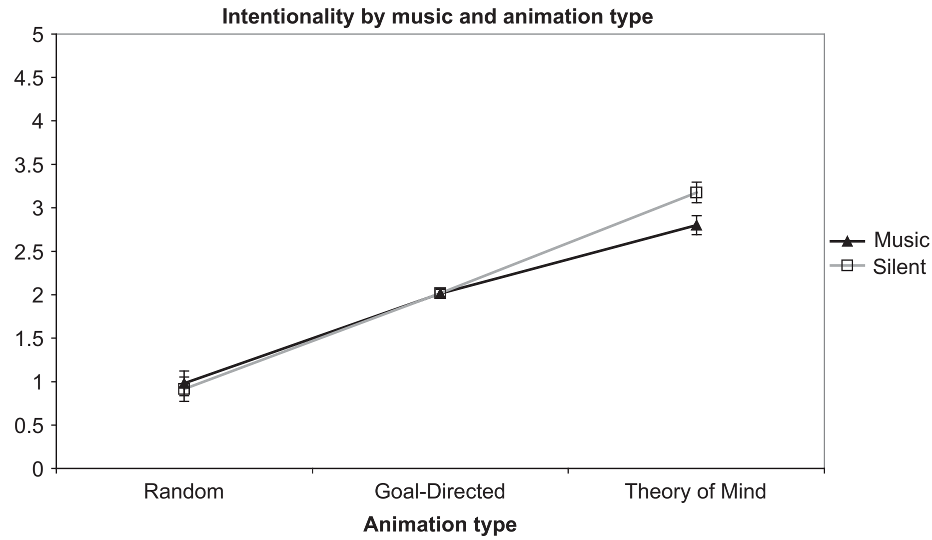

Figure 3 Intentionality ratings of responses of all participants in the music and silent conditions for each animation type.

In summary, the two groups' descriptions did not differ in intentionality for the Rnd and GD animations, but the ASD group used words with less intentionality than the control group for the ToM animations. Intentionality scores also showed a relation to parental report of social behavior, as measured by SRS scores. In addition, all participants' descriptions of the ToM animations contained less intentionality when there was music present. Music did not affect the intentionality of the descriptions for the Rnd or GD animations.

Subgroups within the autism spectrum. Of the 26 participants in the ASD group who were included in all previous analyses, there were 13 children who had been diagnosed with Asperger syndrome (AS) and 10 children who had been diagnosed with PDD-NOS (see Table 3 for descriptive statistics). The remaining three were diagnosed with classical autism. Because this was such a small number, these three were excluded

Table 3 Descriptive Statistics and *t*-tests for Participants with Asperger Syndrome and PDD-NOS

	Age(yr:mo)	FSIQ	VIQ	PIQ
Asperger syndrome (*N* = 13)				
Mean	13:3	108	107	105
SD	1:9	14	15	14
Range	10:11–19:4	86–133	88–132	86–129
PDD-NOS (*N* = 10)				
Mean	13:9	88	81	98
SD	1:6	6	10	8
Range	10:10–16:2	82–100	62–93	87–111
$t(21)$	.62	4.23**	4.91**	1.44

FSIQ: Full scale IQ, VIQ: verbal IQ, PIQ: performance IQ.
$^{**}p < .001$.

from the following analysis only. However, examination of the data from the three participants with classical autism showed the appropriateness and intentionality of their responses to be more similar to that of the PDD-NOS group than that of the AS group.

We compared the music matching, length, appropriateness, and intentionality results from the AS group to those from the PDD-NOS group using repeated measures ANOVAs with subgroup diagnosis and animation as factors for music matching and diagnosis, animation and music condition as factors for the other three variables. There were no significant differences between subgroups in music matching or length.

For appropriateness, there was no main effect of subgroup $F(1, 21) = 1.07$, $p = .31$, but there was a significant interaction between music condition and subgroup diagnosis, $F(1, 21) = 4.34$, $p < .05$. This was due to a reduced appropriateness in the responses from the PDD-NOS group for the silent animations, but no difference in appropriateness for the animations with music.

For intentionality, again subgroup diagnosis was not a significant factor, $F(1, 21) = 1.62$, $p = .22$, but the interaction between diagnosis and music condition approached significance, $F(1, 21) = 3.03$, $p < .10$ (failure to reach significance may be due to low observed power; power = .30). As can be seen in Figure 4a, the TD, AS, and PDD-NOS groups' intentionality responses were similar to each other in the music condition but differed in the silent condition. While the group with PDD-NOS and the TD group showed similar patterns, with their intentionality of response higher in the silent condition than in the music condition, the group with AS showed the opposite pattern. Their intentionality responses were higher for the animations with music than for the silent animations. In addition, the three-way interaction among music condition, animation type, and diagnosis approached significance, $F(2, 42) = 3.09$, $p < .06$, (observed power = .56). When the animation types were examined separately, the different patterns between the AS and PDD-NOS group were only evident for the Rnd animations; the groups did not differ for the GD or ToM animations (see Figure 4b).

Verbal IQ

Because verbal IQ was approximately one *SD* different between groups, we explored its relationship with the dependent variables by using Pearson correlations. We found that, of all the variables, VIQ was only significantly correlated with Appropriateness on the Rnd animations, $r = .32$, $p = .02$. Thus, individuals with higher IQs gave responses that were more appropriate for the Rnd animations. Because this variable showed no main effect or interaction with the group variable, there were no group effects that could be accounted for by VIQ differences. Intentionality was the only variable that showed an impairment in the ASD group, so we further examined this using an ANCOVA with VIQ as the covariate. VIQ did not show any significant between-subjects effects, $F(1, 49) = 0.023$, $p = .9$, nor did it interact with the music/silent or animation factors (all p values > .5). We conclude, therefore, that it is neither a significant factor in the general linear model nor is it a significant predictor.

DISCUSSION

The analysis of both the music and silent conditions yielded similar results to those obtained by Castelli et al. (2002) and Abell et al. (2000). Both of these studies found differences between ASD and control groups in the appropriateness of mentalizing descriptions

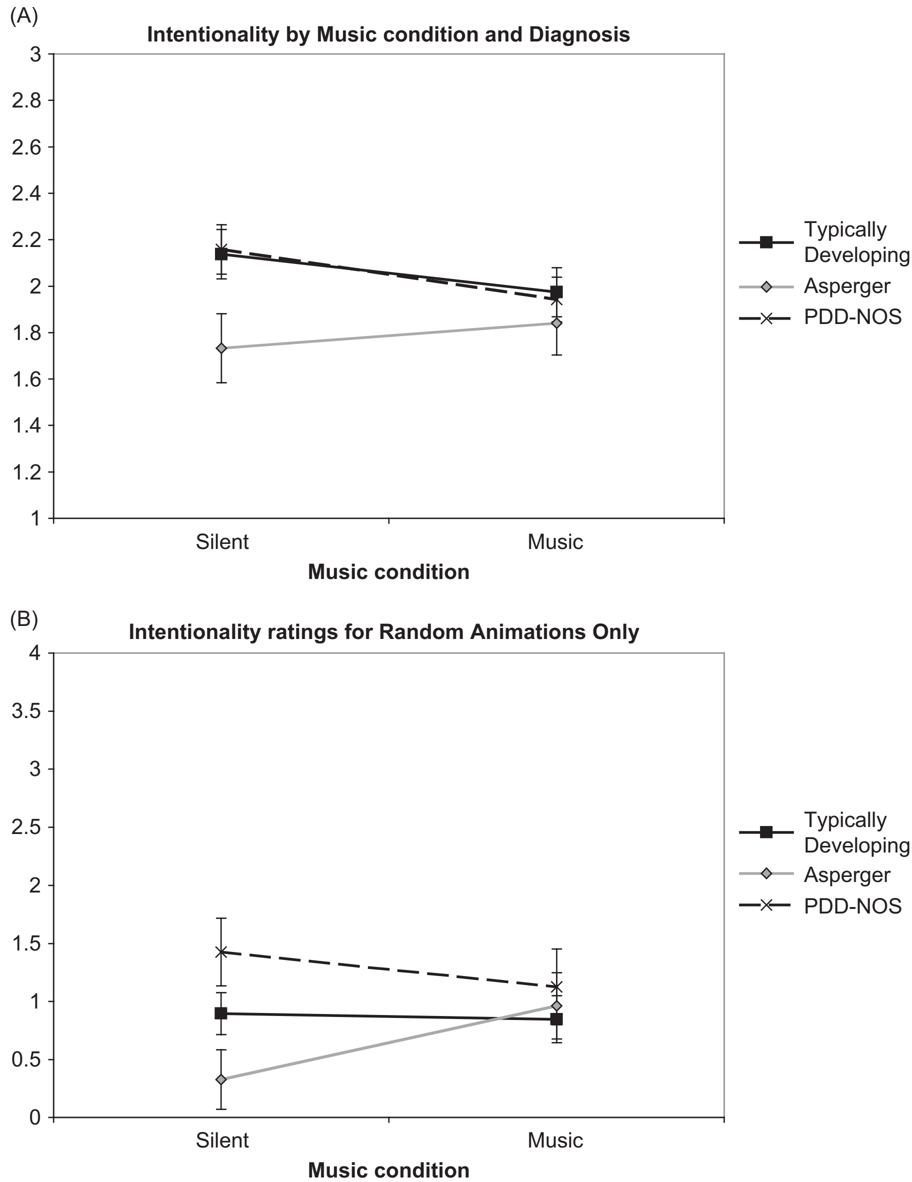

Figure 4 (a) Intentionality ratings by each diagnostic subgroup for the music and silent conditions, collapsed across animation type and (b) intentionality ratings by each diagnostic subgroup for the music and silent conditions for only the random animations.

of ToM animations. We did not find a significant intergroup difference in appropriateness of descriptions, but we did find that the ASD group used fewer “mentalizing” words in their descriptions of the ToM animations than the control group did. The similarity of our findings with these earlier studies validates our inclusion of younger (adolescents rather than adults) participants, drawn from the broader autism spectrum (PDD-NOS).

There is evidence that individuals with ASD are impaired in integrating auditory and visual information at higher levels of processing, including processing of linguistic information (Bebko, Weiss, Demark, & Gomez, 2006; O'Connor, 2007; Smith & Bennetto, 2007). This is relevant in the current context as the task depended, in part, on intact integration of higher level auditory and visual stimuli. Specifically, we found that the responses from the participants with ASD to the question "How well does the music match?" did not differ from those of controls. While previous findings have shown that children with ASD are unimpaired in matching music to static representations (line drawings) of mental states and emotions (Heaton et al., 1999, 2008), this is the first study to show that they do not differ from controls in rating how well music matches a moving picture. Like the controls, the children with ASD gave higher ratings to the animations with more social complexity, possibly because these contained more visual information with which to match the music. In addition, this suggests that all participants were successful at some level in integrating the auditory input of the music with the visual input from the animations. This finding is not consistent with evidence showing difficulties in integrating information across modalities in autism (Bebko et al., 2006; Iarocci & McDonald, 2006; O'Connor, 2007; Smith & Bennetto, 2007); although face/voice integration may rely on different neural mechanisms to animation/music associations, and face/voice integration clearly makes higher social cognitive demands.

In order to provide a clearer interpretation of the findings, the results that were not affected by the presence of music will be discussed first. The results from the length measure show that that all participants gave longer descriptions to the ToM animations than to the other two types of animations. This replicates the Castelli et al. (2000) findings with healthy adults and can be attributed to the increased complexity of the ToM animations relative to GD and Rnd. The results from the appropriateness measures showed that all participants described the ToM animations less appropriately than they described the GD and Rnd animations, that is, their descriptions of the animations were less well matched with what the animation designers intended. As with the increased length of descriptions for the ToM animations, this may be due to the increased complexity of the interactions depicted in the ToM animations. All participants included more intentionality words in their descriptions for the ToM animations than for the GD, and more for the GD than for the Rnd. This is an expected result, as by design ToM animations included higher levels of intentionality (meaning the shapes showed more purposeful movement) than the GD, and the GD included higher levels of intentionality than the Rnd.

The children in the ASD group did not differ from controls in the length of utterances in response to the ToM animations. However, the overall length of utterance for the ASD group collapsed across conditions was longer. This reflects smaller differences across animation types as well as higher length of utterance scores than controls for the Rnd and GD animations. This finding, showing that ASD children did not show selective increases in length of descriptions in response to the ToM animations relative to the other two types may reflect the reduced salience of more "social" interactions.

The results from the appropriateness measure showed that the children with ASD did not differ from the control group.

On the intentionality measure, the significant interaction between diagnosis and animation revealed that the control group's descriptions contained more intentionality words than the ASD group's for the ToM animations, and this was negatively correlated with SRS scores across groups. Thus, the lower an individual's score on the SRS (indicating a greater amount of social responsiveness) the more intentionality in his or her responses for the ToM animations.

In summary, the results of the present study have replicated the main findings of previous studies with the exception of the appropriateness measure. Previous studies have shown participants with ASD to be impaired in appropriateness of description relative to the TD group, but the present study found no difference.

The Effect of Music

On average, all participants' descriptions of animations with music were less appropriate than their descriptions of the silent animations. This suggests that the music provided a conflicting source of information and in many cases altered the participants' perception of the visual events. Indeed, it was noted that all participants exhibited a change in appropriateness as a result of music on at least one of the eight animations, confirming that the music modified the participants' perception of the events in the animation.

In addition, all participants' ToM descriptions were decreased in intentionality by the addition of the music. This could be due to the music being a conflicting information source (providing alternate emotional interpretation of the events of the animations) or being a simple distraction. Below are two examples of participants' descriptions of the same animations ("Mocking") with and without music. In this animation, the blue triangle is following the red triangle and imitating it without the red triangle noticing until the end. See Appendix B for more specifics on scoring.

TD participant

Silent. Ah, the dance . . . Oh . . . it looks like someone following someone else without them noticing. And at the end, he realizes; the other one just runs away. *Scores: Intentionality = 4.5, Appropriateness = 3, Length = 3.5.*

Music. That's it. They were dancing, even though they're triangles.
Scores: Intentionality = 3, Appropriateness = 1, Length = 2.

Participant with ASD

Silent. All right, it looked like one of the triangles was walking along, and then the blue smaller one comes along, and he's following the other one and imitating him. The red one turns around, and the blue one is like "Do do doo," and that happened several times.
Scores: Intentionality = 4.5, Appropriateness = 3, Length = 5.

Music. The music kind of changes the way it seems. Instead of one trying to annoy the other, it looks like this dance where they go, and then they turn, and then they do that, and then one pokes the other by accident when they're walking and the other turns around and says "Hey, why'd you do that?" The one that poked him just runs away and says "bleahh."
Scores: Intentionality = 3, Appropriateness = 1.5, Length = 5.

These are examples of how the addition of music decreased both the appropriateness and intentionality scores of ToM animations. The length score of the participant with ASD did not change in this specific example; though as a group the ASD participants' length of description was decreased by the music, as was the TD group's.

Overall, the most important aspect of the difference between the music and silent conditions is that both groups were influenced in the same direction on all three of the dimensions on which descriptions were scored: length, appropriateness, and intentionality. Additionally, there was no difference between groups in ratings of how well the music matched the animations. The only evidence of a group difference was the trend toward increases rather than decreases in intentionality scores on the random condition for participants with Asperger syndrome compared to those with PDD-NOS and TD. The participants with Asperger syndrome attributed almost no intentionality to the random animations in the silent condition, but in the music condition the intentionality of their descriptions was increased, while the other two groups showed the opposite trend.

CONCLUSION

Individuals with ASD are impaired in attributing intentionality to abstract shapes when the shapes' interactions appear to be socially complex. This result coincides with our first hypothesis that deficits in social attribution cannot be influenced by music, at least in these combinations presented. At the same time, music had no differential effects on the ASD and TD groups. Music altered appropriateness across all participants and animation types and decreased intentionality scores of all participants' responses on the more socially complex animations. We also found evidence suggesting that there are differences between subgroups within the autism spectrum in the effect music has on appropriateness and intentionality of descriptions, but this is preliminary and should be studied further with larger samples of participants. As measured by the effect that music has on their social attribution, adolescents with ASD show no deficits in their ability to integrate music with moving visual displays or their ability to extract meaning from musical excerpts. Future studies are needed to explore this in greater depth, particularly with respect to subgroups within the autism spectrum, but this is evidence that music is indeed a domain of relatively spared cognition for individuals with ASD.

Original manuscript received April 3, 2008
Revised manuscript accepted November 2, 2008
First published online January 13, 2009

REFERENCES

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

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

Balkwill, L. L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, *46*(4), 337–349.

Baron-Cohen, S. (2001). Theory of Mind and autism: A review. In L. M. Glidden (Ed.), *International review of research in mental retardation: Autism* (pp. 169–184). San Diego, CA: Academic Press.

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

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

Bebko, J. M., Weiss, J. A., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry*, *47*(1), 88–98.

Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception*, *18*(4), 427–454.

Boltz, M. G. (2004). The cognitive processing of film and musical soundtracks. *Memory & Cognition*, *32*(7), 1194–1205.

Boltz, M. G., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory & Cognition*, *19*, 593–606.

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

Capps, L., Yirmiya, N., & Sigman, M. (1992). Understanding of simple and complex emotions in non-retarded children with autism. *Journal of Child Psychology and Psychiatry*, *33*(7), 1169–1182.

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

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

Cohen, A. J. (2001). Music as a source of emotion in film. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 249–272). New York: Oxford University Press.

Constantino, J. N., Davis, S., Todd, R., Schindler, M., Gross, M., Brophy, S., et al. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders, 33*, 427–433.

Cross, I. C. (in press). The evolutionary nature of musical meaning. *Musicae Scientiae*.

Cuddy, L. L., Balkwill, L. L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: A study of "tone deafness" among university students. *Annals of the New York Academy of Sciences*, *1060*, 311–324.

Fisher, N., Happé, F., & Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, 46*(4), 409–419.

Fombonne, E. (2005). Epidemiology of autistic disorder and other pervasive developmental disorders. *Journal of Clinical Psychiatry*, *66*(Suppl. 10), 3–8.

Fombonne, E., Zakarian, R., Bennett, A., Meng, L., & McLean-Heywood, D. (2006). Pervasive developmental disorders in Montreal, Quebec, Canada: Prevalence and links with immunizations. *Pediatrics*, *2006*(118), 139–150.

Heaton, P., Allen, R., Williams, K., Cummins, O., & Happé, F. (2008). Do social and cognitive deficits curtail musical understanding? Evidence from autism and Down syndrome. *British Journal of Developmental Psychology*, *26*, 171–182.

Heaton, P., Hermelin, B., & Pring, L. (1999). Can children with autistic spectrum disorders perceive affect in music? An experimental investigation. *Psychological Medicine*, *29*, 1405–1410.

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

Hobson, R., Ouston, J., & Lee, A. (1989). Naming emotion in faces and voices: Abilities and disabilities in autism and mental retardation. *British Journal of Developmental Psychology*, *7*(3), 237–250.

Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders*, *36*(1), 77–90.

Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The social attribution task. *Journal of Child Psychology and Psychiatry*, *41*(7), 831–846.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, *94*, 412–426.

Levitin, D. J., Cole, K., Chiles, M., Lai, Z., Lincoln, A., & Bellugi, U. (2004). Characterizing the musical phenotype in individuals with Williams syndrome. *Child Neuropsychology, 10*, 223–247.

Lipscomb, S. D., & Zehnder, S. M. (2004). Immersion in the virtual environment: The effect of a musical score on the video gaming experience. *Journal of Physiological Anthropology and Applied Human Science*, *23*(6), 337–343.

Marshall, S. K., & Cohen, A. J. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception, 6*(1), 95–112.

O'Connor, K. (2007). Brief report: Impaired identification of discrepancies between expressive faces and voices in adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *37*, 2008–2013.

Ozonoff, S., & Miller, J. N. (1995). Teaching theory of mind: A new approach to social skills training for individuals with autism. *Journal of Autism and Developmental Disorders*, *25*, 415–433.

Rutter, M., Bailey, A., & Lord, C. (2003). *SCQ: Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development*, *77*(4), 1034–1049.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Smith, B. (1995). *Psiexp: An environment for psychoacoustic experimentation using the IRCAM musical workstation*. Paper presented at the Society for Music Perception and Cognition Conference, University of California, Berkeley.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, *48*(8), 813–821.

Sparrow, S., Balla, D., & Dicchetti, D. (1984). *Vineland Adaptive Behavior Scales, Expanded edition*. Circle Pines, MN: American Guidance Service.

Taylor, S. E. (1997). The social being in social psychology. In D. Gilbert, S. Fiske, & G. Lindsey (Eds.), *The handbook of social psychology* (Vol. 4, pp. 58–95). New York: McGraw-Hill.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

Wing, L. (1997). Syndromes of autism and atypical development. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 148–170). New York: John Wiley & Sons, Inc.

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, *9*(1), 11–29.

Ziv, N., & Goshen, M. (2006). The effect of "sad" and "happy" background music on the interpretation of a story in 5 to 6-year-old children. *British Journal of Music Education*, *23*(3), 303–314.

APPENDIX A

1. List of music used

Artist or Composer	Song Title	Album
Carl Stalling	In Session (1951–1956)	The Carl Stalling Project — Music from Warner Bros. Cartoons 1936–1958
Carl Stalling	Powerhouse and other cuts from the early 50s	The Carl Stalling Project — Music from Warner Bros. Cartoons 1936–1958
Carl Orff	Uf dem Anger — Dance	Carmina Burana
Courtney Pine	The 37th Chamber	Modern Day Jazz Stories
David Grisman	Minor Swing	Hot Dawg
Duke Ellington	Creole Love Call	The Essential Duke Ellington
Danny Elfman	End Theme	The Nightmare Before Christmas
Herb Alpert and the Tijuana Brass	Spanish Flea	Greatest Hits
John Coltrane	Giant Steps	Giant Steps
Modest Mussorgsky	The Hut on Fowl's Legs (Baba Yaga)	Modest Mussorgsky: Pictures at an Exhibition

2. Pairing of music with animations, organized by animation type (level of social complexity)

	Random		Goal-Directed		Theory of Mind	
	Animation	Music	Animation	Music	Animation	Music
Silent	Billiard		Leading		Mocking	
	Star		Fighting		Surprise	
			Dancing		Seducing	
Music	Billiard	Giant Steps	Leading	Creole Love Call	Mocking	Minor Swing
	Star	The 37th Chamber	Leading	Nightmare Before Christmas End Theme	Mocking	Powerhouse and other cuts from the early 50's
			Fighting	Spanish Flea	Surprise	Uf dem Anger - Dance
			Dancing	The Hut on Fowl's Legs (Baba Yaga)	Seducing	In Session (1951–1956)

APPENDIX B

ANIMATIONS TASK-SCORING PROCEDURES

Scoring verbal descriptions: The verbal descriptions provided by subjects after each animation are transcribed verbatim and coded in terms of three different dimensions: Intentionality, Appropriateness, and Length. For intentionality ratings, the highest level response that is provided within a particular narrative should be rated. For appropriateness ratings, the overall content of the response is considered in making ratings.

1. Intentionality Score (0 to 5)

The Intentionality score reflects the degree to which the subject describes complex, intentional mental states. Degree of attribution of mental states to the triangles (agents) is calculated by analyzing the content of each narrative provided by the subjects. To minimize the use of subjective interpretations of others' language use, terminology, idioms, and so forth, the analysis focuses *exclusively on the type of verb* that is used to describe the triangles' actions. The degree of intentionality reflected in the verb is rated with a numerical scale that ranges from 0 to 5. Ratings follow an "intentionality ladder," with scores moving upwards in accordance with the degree to which the narrative describes purposeful movement, interaction between agents, and attribution of mental states to one or both agents in the animations.

0 = No purposeful action, no mental state attribution (even if 2 agents are described).

At the bottom of the intentionality ladder, the description does not refer to purposeful movement, interaction between the characters, or attribution of intentionality. Any movement described is random.

Examples: moving around, floating, bouncing off, spinning, orbiting.

1 = Purposeful action but no interaction between agents, no mental-state attribution.

In the next step up in the ladder, the movement of one or both agents is described as having a purpose or goal, but there is no interaction between the agents and no attribution of mental states.

Examples: walking, swimming, running, galloping, jumping.

2 = Purposeful action that involves interaction between agents, but no mental-state attribution.

In this step, the agents are described as moving in a purposeful, interactive manner with each other but there is no attribution of mental states.

Examples: fighting; following, chasing, restraining, guarding, pulling, pushing, wresting, dominating, copying.

3 = Goal-directed intention (i.e., wanting to, trying to) to one agent or an emotion that does not involve reciprocal interaction between agents.

A further step up, the description contains deliberate action with reference to goal-directed, intentional states or emotions of only one agent (or separate attributions for each agent) — mental-state attributions do not occur in the context of reciprocal interaction between the agents. Both goal-directed intentions and emotions are directly observable in the agent's physical motion or physical changes (e.g., sobbing, giggling, shaking in the case of emotions and repeated movement — failing and trying again — towards a target in the case of goal-directed intentions).

Examples: wanting to, trying to, is reluctant, is happy, is angry, subject gives one of the characters a voice ("the little one goes, 'Hey, I don't want to do that'").

4 = Attribution of "transparent" mental states that involve reciprocal interaction between agents.

This step describes deliberate action with reference to complex mental states and emotions that are "transparent" and cannot be inferred directly from observing the agent's clear physical clues. In the case of emotions, this step describes emotions that occur in the context of reciprocal interactions (e.g., emotions of one agent are affected by the behavior of the other agent).

Examples: thinking, being interested, wondering, encouraging, mocking, mimicking, teasing, being friendly with each other, having fun with each other, "the blue one is annoying the red one."

5 = One agent intentionally affecting or manipulating the mental state of another agent.

The highest step describes the most complex types of interactions between two agents on a mental-state level. One agent acts with the goal of affecting, deceiving, or manipulating the other. The description reflects appreciation of how one agent purposely uses awareness of other's mental state to achieve some goal.

Examples: pretending, deceiving, coaxing, surprising, convincing, seducing, tricking, persuading, "the red one probably didn't want to hurt the blue one's feelings," "the red one is trying to make the blue one feel better," "the red one is trying to embarrass or humiliate the blue one."

2. Appropriateness score (0–3)

The Appropriateness score rates how accurately the participant describes the event depicted in the animations, as intended by the underlying scripts for each animation that were developed by the designers of this task. The degree of appropriate description of the animation is determined by analyzing the description of the agents' actions, interactions, and mental states. For example, an appropriate description (score = 3) for the ToM animation in which the big triangle persuades the little one to go out must convey the idea of little triangle's reluctance to go out *and* big triangle's attempts to get the little one out — e.g., "persuading" or "coaxing." A less appropriate description (score = 2) would focus on one aspect of the story or one character only — e.g., little one doesn't want to go out, or, big one is pushing little one to go out. An inappropriate description (score = 1) concerns actions that do not relate to the events or relate to a very minor aspect of the sequence only — e.g., "the two triangles didn't like each other." Finally, when the subject does not provide any description, the narrative is scored 0.

General rules:

3 = Spot-on description of the story or the actions represented. It may be concise, just capturing gist, or discursive.
2 = Partial description of the sequence. Description is related to the sequence but is imprecise or incomplete.
1 = Descriptions that focus solely on a minor aspect of the sequence.
0 = "Don't know" answers.

Random movement sequences:
Billiard (animation with no enclosure)**; Star** (animation with enclosure)

3 = Descriptions implying random or purposeless movement including moving about, bouncing off the walls, or dancing as in dancing lights.
2 = Purposeful movement without interaction, including turning round and getting dizzy, or dancing in a circle.
1 = Purposeful movement implying interaction between the triangles including copying each other.

Goal-directed movement sequences:
Fighting: (animation with no enclosure)

3 = Action implying physical fight (e.g., bashing each other).
2 = Action that conveys the idea of a conflict but is either too specific or too vague (e.g., biting, pushing).
1 = Action that does not relate to conflict (e.g., following each other) or focuses solely on a minor aspect of the sequence.

Leading: (animation with enclosure)

3 = Description that conveys the idea of one leading the other or one following the other.
2 = Description that is related to but somewhat remote from following (e.g., copying, chasing).
1 = Action that does not relate to following/leading or focuses solely on a minor aspect of the sequence.

Dancing: (animation with no enclosure)

3 = Description that conveys the idea of moving in formation (e.g., dancing, making a pattern).
2 = Description that is partially correct or related to dancing (e.g., doing different things — one went one way — the other went the other way).
1 = Action that is not related to dancing (e.g., galloping along or focuses solely on a minor aspect of the sequence).

Theory of Mind movement sequences:
Surprising: (animation with enclosure)

3 = Any mention of tricking, surprising, hiding, hide and seek.
2 = Description that gives part of the story but misses the critical point (see above).
1 = Description not related to any of the events in the sequence or focuses solely on a minor part of action (e.g., knocking on the door).

Mocking: (animation with no enclosure)

3 = Description that conveys idea little triangle is copying big one with the intention of not being noticed (e.g., pretending, hiding, being naughty).
2 = Partially correct description (e.g., following, pursuing, copying).
1 = Description that does not relate to the events (e.g., big triangle not interested) or relates to a very minor aspect of the sequence only (e.g., little triangle ran away).

Seducing: (animation with enclosure)

3 = Description that conveys the little triangle is trapped in and escapes by persuading, tricking the big one (e.g., Little convinces in a seductive way to let him out).
2 = Partial story with minimal action for each character (e.g., Little trying to escape).
1 = Description that is too minimal (e.g., she got out, or unrelated to the sequence).

3. Length Score (0–5) [adapted for use in this study]

0 = no response
1 = one clause
2 = two clauses
3 = three clauses
4 = four clauses
5 = five or more clauses