

Published in final edited form as:

Neuroscientist. 2012 April ; 18(2): 108–113. doi:10.1177/1073858410397208.

Spontaneous theory of mind and its absence in autism spectrum disorders

Atsushi Senju

Centre for Brain and Cognitive Development, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK

Abstract

Theory of mind, the cognitive capacity to infer others' mental states, is crucial for the development of social communication. The impairment of theory of mind may relate to autism spectrum disorder (ASD), which is characterised by profound difficulties in social interaction and communication. In the current article, I summarize recent updates in theory of mind research utilizing the spontaneous false belief test, which assesses participants' spontaneous tendency to attribute belief status to others. These studies reveal that young infants pass the spontaneous false belief test well before they can pass the same task when explicitly asked to answer. By contrast, high-functioning adults with ASD, who can easily pass the false belief task when explicitly asked to, do not show spontaneous false belief attribution. These findings suggest that the capacity for theory of mind develops much earlier than was previously thought, and the absence of spontaneous theory of mind may relate to impairment in social interaction and communication found in ASD.

Imagine the following situation. You happened to come home earlier than usual. You were very hungry, and remembered that your partner keeps her (or his) precious chocolate in the cupboard, which she only eats after unusually hard day's work as a special treat. You know how important the chocolate is to her, but you were so hungry that you took the whole box out of the cupboard and ate about half of it. When she suddenly arrived home, you just had time to put the box under the coffee table before she came in to the living room. She then said that "I've had a really hectic day - I'm exhausted! I think I deserve some chocolate tonight." What would you do?

Most of you would predict that she would get to the cupboard (so you have to do something quickly, before she opens it.). At the same time, you may not realize what a complex and sophisticated reasoning you've just made to generate this prediction, as it would have occurred to you naturally and effortlessly. The reasoning you have just made could be broken down into understanding that she will open the cupboard because (1) she wants the chocolate and (2) she believes that it is still in the cupboard because (3) she doesn't know that you've moved it. Such reasoning is called the theory of mind. It involves inferring others' behaviour based on their mental states, which are opaque and impossible to observe directly. Since the concept of theory of mind was first introduced (Premack and Woodruff, 1978), it has been central to investigations within various fields of cognitive science, including developmental psychology, comparative psychology, cognitive neuroscience and child psychiatry. Theory of mind is essential for human social interaction and communication, and also plays a significant role in human cooperation and moral reasoning (e.g. Waytz, Gray, Epley, and Wegner, 2010). Thus, theory of mind is a cornerstone of human adaptation to complex and effective social institutions.

In the above example, the prediction of the behaviour is based on her false belief, because the belief (i.e., the chocolate is still in the cupboard) is *counterfactual*, or incongruent to the

current state of the world (i.e., the chocolate is actually under the coffee table). Attribution of false belief is the hallmark of theory of mind, because it generates unique predictions of others' behaviour which are impossible solely from the actual states of the world (Dennett, 1978). Thus, many psychologists have treated false belief understanding as "the litmus paper" of theory of mind, and used the so-called false belief test to assess the capacity for theory of mind at various ages and at various psychiatric and/or neuropsychological conditions. In a standard false belief test, an experimenter presents an event in which an object is moved during the absence of a protagonist, and then asks the participant where the protagonist will look for the object. Participants pass the test if they correctly answer that the protagonist will look for the original location of the object, and fail if they choose the current location of the object.

In a landmark study, Baron-Cohen, Leslie and Frith (1985) reported that children with autism, who suffer from severe impairment in social interaction and communication, do not pass the false belief test. This seminal study inspired many follow-up studies, both in typically developing children and children with autism spectrum disorders (ASD). The accumulation of the studies supports the proposal that children with ASD show atypical development of the capacity for theory of mind. The most consistent finding is that before the verbal mental age of 11 years, children with ASD do not pass various versions of the false belief test. Whereas typical 4-year-olds correctly anticipate others' behaviour based on the attribution of false belief, children with ASD, at the same mental age or even higher, incorrectly predict behaviour based on reality without taking the other person's epistemic states (e.g. knowledge and belief) into account (Happé, 1995, Figure 1). This result has been interpreted as evidence that children with ASD (with verbal skills equivalent to typically developing children of under 11 years) fail to represent others' epistemic mental states, or at least fail to do so when others' mental states are different from the child's own. Based on these findings, some scientists suggested that the difficulty in social interaction and communication in ASD derives from impairment in theory of mind (Mindblindness theory, Baron-Cohen, 1995).

However, there are two major problems in the standard false belief test. Firstly, it is a cognitively demanding test. To pass the test, a child has to remember the sequence of the presented event, correctly understand the experimenter's question, and inhibit their prepotent response to answer the actual location of object (Birch and Bloom, 2003). Thus, it is possible that despite having the capacity to represent another person's mental state, children fail the standard false belief test because of a weaker cognitive control or difficulty in pragmatic understanding. Secondly, as can be seen from Figure 1, individuals with ASD with higher verbal skills do pass the standard false belief test.

However, the qualitative difficulties in social interaction and communication persist even in these "high-functioning" individuals with ASD. For example, in an experimental setting, they still show subtle difficulties in understanding non-literal utterance (Happé, 1994) and faux-pas (Baron-Cohen, O'Riordan, Stone, Jones, and Plaisted, 1999), and fail to correctly infer complex mental states from photographs of eyes (Baron-Cohen, Jolliffe, Mortimore, and Robertson, 1997) and from cartoon animations (Castelli, Frith, Happé, and Frith, 2002). The capacity to correctly represent another person's false belief may not be sufficient for a fully-functional theory of mind, which works in various contexts.

Young infants spontaneously attribute false belief to others

Infant scientists have revealed that infants have amazing cognitive skills well before they begin to walk or speak. For example, even newborns discriminate faces from non-facial objects, and differentiate the direction of eye gaze of faces. Within the first year of life,

infants show a rudimentary understanding of physical principles such as rigidity and gravity, as well as an understanding of others' minds in terms of concepts such as goals and perception (e.g. Spelke and Kinzler, 2007).

As infants cannot answer verbal questions, infant scientists rely on infants' spontaneous responses to carefully controlled stimuli to assess these capabilities. The majority of studies make inferences about infants' cognitive operation from the pattern of their looking behaviour. Preferential looking techniques, for example, involve presenting two stimuli side-by-side, to see if the infant prefers looking at one over the other. A preference suggests that infants can discriminate between the two stimuli. More recently, infant scientists have begun to use eye-tracking devices, which allow them to record exactly which *area* of the stimuli infants are watching.

In the last 5 years, such studies of infants' spontaneous behaviour have revealed that children pass a false belief test well before the age of 3. Take one of our previous studies using eye-tracking techniques, for example (Southgate, Senju, and Csibra, 2007, Figure 2). In this study, infants initially watched familiarization movies, in which a puppet hid a toy in one of two boxes (Figure 2A). Then, both windows were briefly illuminated and a chime sounded (Figure 2B). About 2 seconds after the illumination and the chime, the actor would open one of the two windows, reach out of it and take the toy (Figure 2C). After a few trials, infants would learn the contingency between the cues (i.e. the illumination and the chime) and the actor's subsequent action, anticipate the action after the cues, and fixate on the correct window (i.e. the window closer to the box in which the toy was hidden) *before* the actor opened the window. Once this anticipatory look was established, we then presented a false belief movie, in which the puppet took the toy from the box and disappeared while the actor was looking away from the stage (Figure 2D). When the actor turned back, the cues were presented. In this experiment, the anticipatory looking of 24-month-old infants showed that they anticipated that the actor would reach for the box in which she had last seen that the toy was hidden. The results clearly demonstrated that these infants attributed false belief to the actor, and predicted the actor's behaviour based on her belief, not the actual state of the world.

To date, a number of studies, including ours described above, have demonstrated that typically developing infants pass a spontaneous false belief test within the second year of life. The majority of these studies measured infants' spontaneous behaviour such as looking time or eye-tracking (Baillargeon, Scott, and He, 2010). Others used more active interaction with the experimenter such as word learning (Southgate, Chevallier, and Csibra, 2010) or helping behaviour (Buttelmann, Carpenter, and Tomasello, 2009).

These findings confirm the claim that the standard explicit false belief test is hampered by high cognitive and verbal demands, which preclude us from revealing the capacity for false belief attribution in younger infants. It is not clear why young infants have difficulty in standard false belief tests, but some scientists argue that it could be due to immature pragmatic skills. For example it is possible that younger infants might prematurely interpret the "Where" question as referring to the actual location of the object (Csibra and Southgate, 2006). Others argue that the main difficulty is in selecting an appropriate answer from other, possibly more salient, alternatives such as the actual location of the object (Baillargeon and others., 2010). Note that these possibilities are not mutually exclusive, and further studies will be required to understand the developmental course of spontaneous and explicit false belief attributions.

Absence of spontaneous false belief attribution in individuals with ASD

Spontaneous theory of mind is critical not only for infants, but also for adults. Unlike experiments, the real social world is fluid and rapidly changing. We have to process socially relevant information rapidly, spontaneously and on-line, in order to achieve day-to-day social interaction. Let's return to the hypothetical social situation described at the beginning of this article. To achieve good social interaction with your partner, you have to detect all the relevant information (e.g. her facial expression, gaze and utterances), predict her behaviour (e.g. looking for the chocolate in the cupboard), and act quickly. All these cognitive operations need to be conducted spontaneously, as in daily social interaction it is highly unlikely that you will be offered explicit information about another person's mental state. For example, you wouldn't expect your partner to ask "Where do you think I will look for the chocolate?". Thus the absence of spontaneous theory of mind would cause difficulty in social interaction and communication, even in adults with high verbal and cognitive skills.

We hypothesized that individuals with high-functioning autism spectrum disorder, even though they can easily pass standard false belief tests (Figure 1), do not pass the spontaneous false belief test. To test this hypothesis, we used the same experiment as with 2-year-old infants on adults with Asperger Syndrome. Asperger Syndrome is a sub-category of ASD, with similar symptoms but no developmental delay in verbal skills (American Psychiatric Association, 2000). In our experiment (Senju, Southgate, White, and Frith, 2009), 19 adults with Asperger Syndrome watched the video stimuli we used in Southgate and others (2007), and their eye movements were recorded with an eye-tracker. No instruction was given to the participants. For comparison with typical development, we also recruited 17 neurotypical adults with similar age, gender ratio and general cognitive skills. All participants in both groups passed standard false belief tests. All participants also showed anticipatory looking by the end of familiarization trials, suggesting that adults with Asperger Syndrome, as well as neurotypical adults, spontaneously anticipate another person's action based on her goal.

However, the test trials showed a very different picture. The Asperger group showed significantly less anticipatory looking toward the correct side than the neurotypical controls did. A follow-up analysis revealed that the anticipatory looking of neurotypical controls were significantly biased to the correct location. This is not the case for Asperger group, whose anticipatory looking behaviour was not biased to either side (Figure 3).

Our study demonstrates that adults with Asperger syndrome do not spontaneously anticipate others' actions in a nonverbal task, closely modelled on the standard false belief test which they pass with ease. In particular, the contrast with neurotypical 2-year-olds who showed spontaneous looking to the correct location on the same task (Southgate and others, 2007) is quite notable. It is unlikely that the general lack of motivation is to blame, as all the participants in the Asperger group showed correct anticipatory looking in the familiarization trials, in which false belief attribution is not necessary.

The results of this study are consistent with the clinical profile of high-functioning individuals with ASD, who show difficulties in social communication in real life despite performing fairly well in a well-controlled experimental or training context. It is also consistent with the findings that training on false belief tests does not necessarily improve social adaptation in ASD (Ozonoff and Miller, 1995): the capacity for false belief attribution may not be sufficient to deal with its spontaneous use in a fluid and rapidly changing "real" social world.

Why do individuals with ASD fail to develop spontaneous theory of mind? There are at least two possibilities. Firstly, they may have a modular impairment in theory of mind (Leslie,

Friedman, and German, 2004), and thus have to develop compensatory strategies to infer other people's mental states. Such a compensatory strategy may not be efficient enough to infer mental states spontaneously and effortlessly in the fluid and rapidly-changing social environment. Secondly, it is possible that socially relevant cues, such as another person's gaze direction or emotional expression, do not spontaneously recruit social cognition in ASD. For example, a recent study demonstrated that infants at high risk to develop ASD showed a slower and less organized cortical response to others' gaze direction (Elsabbagh and others, 2009), even though they have no difficulty in spontaneously orienting to another person's eyes (Chawarska and Shic, 2009). Note that these two possibilities are not mutually exclusive because an early impairment in processing socially relevant information may lead to an atypical development of theory of mind (Senju and Johnson, 2009). Further developmental studies will be necessary to understand the developmental trajectory of spontaneous and deliberate theory of mind, both in typically developing individuals as well as in individuals with ASD.

Future directions

Recent developments in theory of mind research have revealed that young infants spontaneously attribute false beliefs to others well before they can pass explicit false belief tests. By contrast, high-functioning adults with ASD, who easily pass explicit false belief tests, do not show spontaneous false belief attribution. These new findings change our understanding of the development of theory of mind both in typical development and in individuals with ASD. However, as is often the case with such new findings, it creates more questions than answers.

Firstly, we need to identify the mechanisms underlying spontaneous false belief attribution. Note that it is not unnecessary automatic or obligatory. For example, some studies (Apperly, Riggs, Simpson, Chiavarino, and Samson, 2006; Lin, Keysar, and Epley, 2010) have demonstrated that even in typically developed adults, belief attribution can be easily disrupted by task demands. Thus we need to understand how and when the theory of mind is computed spontaneously in real social interaction, what kind of social and/or environmental cues prompt such computation, and how it underlies successful social communication. Further investigation is also required for the neural substrates of spontaneous theory of mind, especially how far it does or does not overlap with the regions related to the processing of theory of mind in response to explicit instruction.

Secondly, we need to revisit the developmental course of theory of mind. Recent studies consistently show that infants pass spontaneous false belief tests by about 15-18 months of age (Baillargeon and others, 2010). Can younger infants pass the task? If not, is it because of the development of theory of mind itself, or due to other task-relevant skills such as working memory, oculomotor control and/or motivation? Are there any environmental factors, such as the number of older siblings or the mode of parent-infant communication, which would affect the development of spontaneous theory of mind? Most of these questions have been asked for the development of explicit false belief attribution, but we would have to begin again with this new theory of mind test at hand. Another important question is, as I discussed above, to investigate why young toddlers fail in explicit false belief test even though they can pass spontaneous theory of mind tests.

Finally, the question remains why adults with ASD in our study show an absence of spontaneous false belief attribution, even though they have no difficulties in explicit false belief reasoning, and how it relates to their impairment in social communication. For example, is there any way we can prompt theory of mind computation in individuals with

ASD other than explicit verbal instruction? Can we train individuals with ASD to spontaneously mind-read? If so, does it change their social adaptation in real life?

Theory of mind has been a core topic in various fields of science which try to understand the biological basis of human social behaviour. Recent updates in theory of mind research brought about by the introduction of the spontaneous theory of mind test has changed our views about the development of theory of mind, and the nature of its impairment in ASD. Future studies on this topic will help us understand how humans process mental states in a fluid and rapidly changing social world, how this helps establish successful human social communication and its development, and how its impairment relates to difficulties in social adaptation in ASD.

Acknowledgments

Grant support

UK Economic and Social Research Council Research Fellowship (RES-063–27–0207)

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed., text revision, DSM-IV-TR. Washington, DC: 2000.
- Apperly IA, Riggs KJ, Simpson A, Chiavarino C, Samson D. Is Belief Reasoning Automatic? *Psychological Science*. 2006; 17(10):841–844. [PubMed: 17100782]
- Baillargeon R, Scott RM, He Z. False-belief understanding in infants. *Trends in Cognitive Sciences*. 2010; 14(3):110–118. [PubMed: 20106714]
- Baron-Cohen, S. *Mindblindness : An essay on autism and theory of mind*. MIT Press; Cambridge, MA: 1995.
- Baron-Cohen S, Jolliffe T, Mortimore C, Robertson M. Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry*. 1997; 38(7):813–822. [PubMed: 9363580]
- Baron-Cohen S, Leslie AM, Frith U. Does the autistic child have a “theory of mind”? *Cognition*. 1985; 21(1):37–46. [PubMed: 2934210]
- Baron-Cohen S, O’Riordan M, Stone V, Jones R, Plaisted K. Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*. 1999; 29(5):407–418. [PubMed: 10587887]
- Birch SA, Bloom P. Children are cursed: an asymmetric bias in mental-state attribution. *Psychological Science*. 2003; 14(3):283–286. [PubMed: 12741755]
- Buttelmann D, Carpenter M, Tomasello M. Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*. 2009; 112(2):337–342. [PubMed: 19524885]
- Castelli F, Frith C, Happé F, Frith U. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*. 2002; 125(Pt 8):1839–1849. [PubMed: 12135974]
- Chawarska K, Shic F. Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year old children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*. 2009; 39(12):1663–1672. [PubMed: 19590943]
- Csibra G, Southgate V. Evidence for infants’ understanding of false beliefs should not be dismissed. *Trends in Cognitive Sciences*. 2006; 10(1):4–5. [PubMed: 16318923]
- Dennett D. Cognition and consciousness in nonhuman species - comment. *Behavioral and Brain Sciences*. 1978; 1(4):568–570.
- Elsabbagh M, Volein A, Csibra G, Holmboe K, Garwood H, Tucker L. Neural correlates of eye gaze processing in the infant broader autism phenotype. *Biological Psychiatry*. 2009; 65(1):31–38. others. [PubMed: 19064038]

- Happé FG. An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*. 1994; 24(2):129–154. [PubMed: 8040158]
- Happé FG. The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*. 1995; 66(3):843–855. [PubMed: 7789204]
- Leslie AM, Friedman O, German TP. Core mechanisms in “theory of mind”. *Trends in Cognitive Science*. 2004; 8(12):528–533.
- Lin S, Keysar B, Epley N. Reflexively Mindblind: Using Theory of Mind To Interpret Behavior Requires Effortful Attention. *Journal of Experimental Social Psychology*. 2010; 46(3):551–556.
- Ozonoff S, Miller JN. Teaching theory of mind: a new approach to social skills training for individuals with autism. *Journal of Autism and Developmental Disorders*. 1995; 25(4):415–433. [PubMed: 7592252]
- Premack D, Woodruff G. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*. 1978; 1(4):515–526.
- Senju A, Johnson MH. Atypical eye contact in autism: Models, mechanisms and development. *Neuroscience & Biobehavioral Reviews*. 2009; 33(8):1204–1214. [PubMed: 19538990]
- Senju A, Southgate V, White S, Frith U. Mindblind Eyes: An Absence of Spontaneous Theory of Mind in Asperger Syndrome. *Science*. 2009; 325(5942):883–885. [PubMed: 19608858]
- Southgate V, Chevallier C, Csibra G. Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*. 2010; 13(6):907–912. [PubMed: 20977561]
- Southgate V, Senju A, Csibra G. Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*. 2007; 18(7):587–592. [PubMed: 17614866]
- Spelke ES, Kinzler KD. Core knowledge. *Developmental Science*. 2007; 10(1):89–96. [PubMed: 17181705]
- Waytz A, Gray K, Epley N, Wegner DM. Causes and consequences of mind perception. *Trends in Cognitive Sciences*. 2010; 14(8):383–388. [PubMed: 20579932]

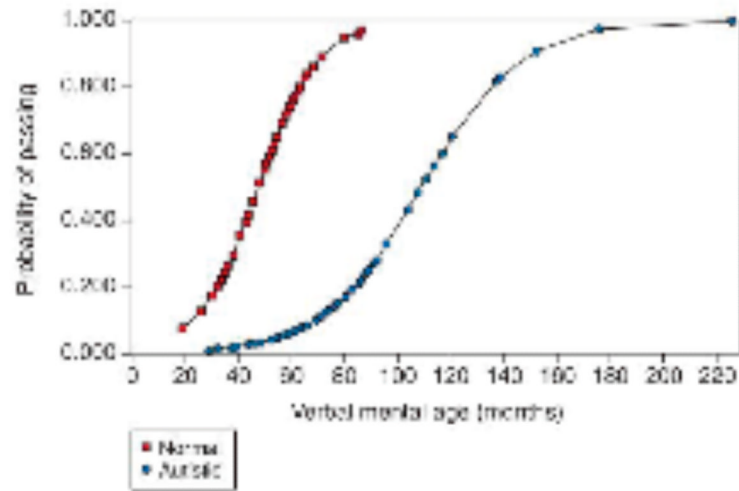


Figure 1. Predicted probability of passing standard false belief tests by verbal mental age. Open square: Individuals with ASD; Closed circle: Typically developing individuals. Reproduced, with permission, from (Happé, 1995)

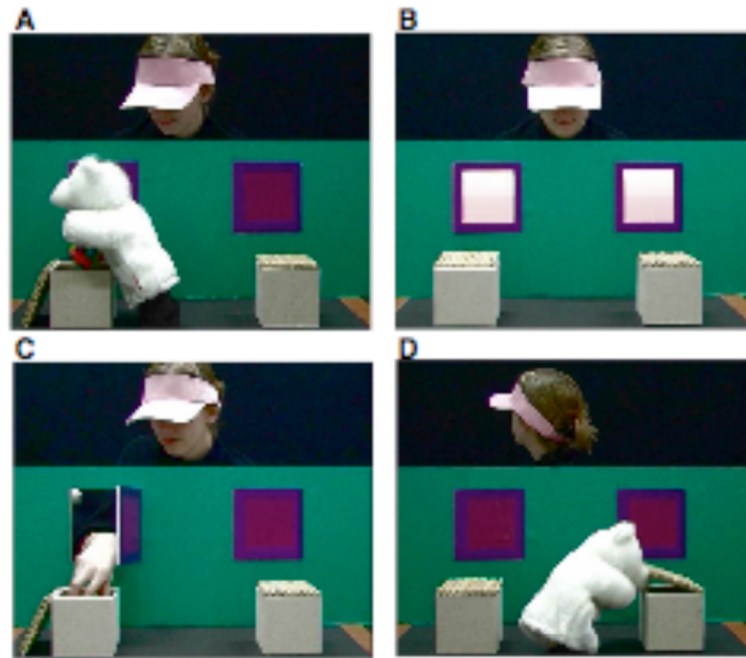


Figure 2. Selected scenes from stimulus movies. In familiarization trials, participants were familiarized to an event in which (**A**) the puppet placed a ball in one of two boxes (**B**) both windows were illuminated and a chime sounded, and (**C**) an actor reached through the window above the box in which the ball was placed, and retrieved the ball. The participants were familiarized to the contingency between (**B**) and (**C**). In (**D**), the puppet moves the ball while the actor is looking away. This operation induces a false belief in the actor about the location of the ball. Reproduced, with permission, from (Senju and others, 2009)

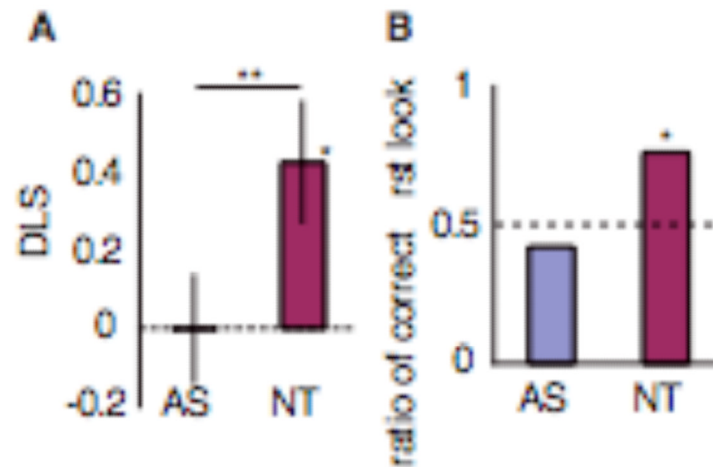


Figure 3.

(A) Mean (\pm SEM) difference looking scores (DLS) and (B) the ratio of the number of participants who made correct first saccades in each group. DLS is calculated by subtracting the looking time for the incorrect window from correct window, and then dividing the difference by the total of the looking times in both correct and incorrect windows. AS: participants with Asperger Syndrome ($N=19$), NT: neurotypical participants ($N=17$), *: $p < .05$, **: $p < .01$, dotted lines: chance level, statistical test used: (A) t-test and (B) binominal test. Reproduced, with permission, from (Senju and others, 2009).