

What Difference Does It Make? Implicit, Explicit and Complex Social Cognition in Autism Spectrum Disorders

Ulrich M. Schaller¹ · Reinhold Rauh¹

Published online: 12 January 2017
© Springer Science+Business Media New York 2017

Abstract We tested social cognition abilities of adolescents with autism spectrum disorders (ASD) and neuro-typically developed peers (NTD). A multi-faceted test-battery including facial emotion categorization (FEC), classical false belief tasks (FBT), and complex social cognition (SC), yielded significantly lower accuracy rates for FEC and complex SC tasks in ASD, but no significant differences in performance concerning FBT. A significant correlation between age and performance in a FEC task and in a complex task was found only in ASD. We propose that dynamic and/or fragmented FEC tasks can elicit deficits in implicit processing of facial emotion more efficiently. The difficulties of ASD in solving complex SC tasks can be ascribed to deficits in the acquisition and application of social schemata.

Keywords Autism spectrum disorder · Social cognition · Theory of mind · Implicit emotion processing · Explicit emotion processing · Social schema

Introduction

At the beginning of our infancy, we as human beings are physiologically premature and deficient (Gehlen 2014; Portmann 1941), and in contrast to many nonhuman animals, we are highly dependent on intense care and social bonding to our loved ones. Embedded in a sociocultural environment, many developmental processes rely on social interaction (Portmann 1941) to the extent that the development of complex human cognition builds upon basic social cognitive processes formed in early childhood (e.g., Rochat et al. 1999; Soto-Icaza et al. 2015). Social cognition (SC) can be regarded as a dynamic constructive and conceiving process of perception, categorization, recall, and evaluation of social stimuli. Experience with the social world emerges from the perception and categorization of behavioral events that unfold in a mutual interplay between self and others over time (Macrae and Miles 2012; Vogeley 2009), and develops by increasing implicit and explicit knowledge of new concepts (Low and Perner 2012). SC acts as an umbrella term and, in this context, we can differentiate several subfunctions like empathy, imitation and social meta-cognition. In this paper we will especially focus on several aspects of mindreading, since mindreading serves as a relevant concept to describe the process of perspective taking and tracking the intentional states of others.

Agency, intentionality, and purposive explanations as efficacious factors of human SC enable us to ascribe observable actions of others to their unobservable mental states. The ability to hypothesize about feelings, desires, intentions and beliefs of oneself and of other people is subsumed in the technical term *Theory of Mind* (e.g., Perner and Lang 1999), and depends on basal processes like (self-)representation, (self-)perception, and categorization. These processes do not always involve a conscious

Electronic supplementary material The online version of this article (doi:10.1007/s10803-016-3008-x) contains supplementary material, which is available to authorized users.

✉ Ulrich M. Schaller
ulrich.schaller@uniklinik-freiburg.de

¹ Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Hauptstrasse 8, 79104 Freiburg, Germany

subjective experience (Brosch et al. 2010) and several studies have shown that judgments within the scope of SC take place without perceptual awareness (Stewart et al. 2012; Todorov et al. 2009). One possibility for defining this fragmentation as a testable construct is to subdivide Theory of Mind (ToM) into implicit and explicit components to describe different aspects of processing social stimuli (Frith and Frith 2012). The implicit component is delineated as unconscious, procedural, non-conceptual, automatic, and spontaneous processing, while explicit processes are described as conscious, declarative, conceptual, and controlled (Kahneman 2011; Low and Perner 2012; Payne 2012). Explicit processes interfere with currently ongoing activity and interrupt automatic and procedural behavior or cognition (Kahneman and Frederick 2002). Implicit processes are characterized by the near impossibility to report how a decision or an action has been accomplished. However, a wide range of cognitive abilities thought to be explicit are based on implicit processes (Gasper and Clore 2000; Higgins 1998; Oikawa 2010).

Contrary to a common belief that the causes for behavior in social situations can be found in conscious decisions, many noteworthy studies (Asch 1951; Festinger and Carlsmith 1959; Milgram 1963; Schachter and Singer 1962) have shown that social actions are more influenced by automatic, unconscious, and uncontrolled processes than expected (Dijksterhuis et al. 2006; Frith and Frith 2012; Payne 2012). Facial expressions are one of the most important carriers of conveying information about the emotional state of the performer (Martinez and Du 2012) and the attribution of emotions to others is mainly based on their facial expressions. Therefore “emotion reading is a fundamental kind of mindreading” (Goldman 2006, p. 113). Generally, neurotypical individuals are highly motivated to turn to and learn to discriminate this continuous sequence of varying muscular contractions and relaxations as categories of basic emotions *ab initio* (Farroni et al. 2004; Grossmann et al. 2008; Johnson et al. 1991; Rigato et al. 2011).

The amount of incoming information by a facial emotional expression requires the ability to reduce the complexity of the continuous sequence and interaction of action units with varying intensity of muscular contractions and relaxations (Brosch et al. 2010). In order to do so the categorization of facial expressions depends on earliest implicit learning as “the acquisition of knowledge that takes place largely independently of conscious attempts to learn and largely in the absence of explicit knowledge about what was acquired” (Reber 1993, p. 5). Referring to the demand reconstruction of the concept of ToM according to Schaafsma et al. (2015), we understand emotion processing in terms of inferring emotional cues from human actions, animate motions and from the facial expressions clearly as a relevant aspect of ToM.

Nevertheless the attribution of FEC to either a lower order process or a higher order feature concerning social cognition is still contested. There are neuroimaging studies suggesting that basic visual perception in NTD (Yamasaki et al. 2002) and ASD (Dawson et al. 2004) is based on lower order processes. On the other hand, there is evidence from current studies and theoretical approaches that perception and categorization of complex emotional stimuli is influenced by two different processes. For the categorization of simplest and basic emotions it is sufficient to process the stimuli by use of purely perceptual automatic processes. With increasing complexity of facial expressions and their embedding in complex social situations the categorization is executed by a cognitive process that uses already acquired background knowledge to conceptualize and categorize the emotions in accordance with the situation (Marchi and Newen 2015). The conceptualizations considered as higher-level cognitive states then influence the perceptual experience (Lindquist et al. 2014).

Although typically developed children are able to name facial basic emotions, the procedure of perception and categorization remains unconscious, automatic, procedural and accordingly implicit (Goldman 2006; Kahneman 2011; Shi et al. 2013). Thus, perceived facial emotions have an important influence on decision making, especially in complex, highly contextual social situations (Van’t Wout and Sanfey 2008; Willis and Todorov 2006; Winston et al. 2002), that take place under time pressure, with limited information, and limited cognitive resources. Some authors describe this kind of social information processing as simple and frugal heuristics in a social world (Hertwig and Hoffrage 2013), characterized as “a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally and/or accurately than more complex methods” (Gigerenzer and Gaissmaier 2011, p. 454). Especially non-explicit, automatically processed facial identity and facial expression are influencing our socially relevant cognitions and decisions (Mehu et al. 2007; Scharlemann et al. 2001; Van’t Wout and Sanfey 2008).

Implicit ToM (iToM) in this context includes the ability to perceive and recognize expressed basic facial emotions (Esteves and Ohman 1993; Todorov 2012; Uleman et al. 2008; Whalen et al. 1998). It can be assessed by tasks concerning the perception and categorization of facially expressed emotions, while explicit ToM (eTOM) is mainly measured by tasks concerning classical false belief tasks (FBT) of first and second order.

Transferring dual process theories of social cognition (for an overview of non-ToM-related accounts in SC, see Sherman et al. 2014) to the domain of ToM, it is conceivable that the perception and categorization of facial expression continuously influences our presumptions, beliefs, and decisions in social situations without our awareness.

Apperly and Butterfill (2009) describe a similar approach: They assume that human infants and non-human animals have access to an implicit ToM system with high cognitive efficiency but low flexibility. During the course of development human children come up gradually with more complex psychological concepts (Apperly and Robinson 2003), that improve with explicit training and build a second ToM system that is flexible but slow and inefficient. NTD adults are equipped with both systems working in parallel.

While it is common in many publications to label tasks using verbal responses as ‘explicit’ and tasks measuring non-verbal behavior as ‘implicit’ (e.g., Kliemann et al. 2013) we start from the premise that predominantly automatic processes are also implicit even if the responses are given verbally. With respect to this issue we are in line with current studies establishing evidence that explicit mentalizing does not influence automatic processing (Schneider et al. 2012, 2014).

ASD, conceived as one of many neurodevelopmental disorders (DSM-5, American Psychiatric Association 2013), is associated with severe and pervasive impairments in social interaction and communication (Frith and Frith 2012; Hill et al. 2004). Some studies suggest, that individuals with ASD and typical IQ are able to solve explicit ToM-tasks (Happé 1995; Senju 2012), likely using compensatory strategies (Frith and Frith 2012; Senju 2012). However, they still score significantly lower in tests concerning complex social situations (Dziobek et al. 2006) and in facial emotion categorization (FEC) tasks (Schuwerk et al. 2014).

As Uljarevic and Hamilton (2012) and Lozier et al. (2014) outlined in their meta-analyses, a considerable emotion recognition deficit forms part of ASD, interpretable with due care because of a presumed influential publication bias on the one hand and most likely underestimated by the use of predominantly static stimuli and lacking time limitations on stimulus presentation on the other. This may have led to similar accuracy rates between neurotypicals (NTD) and ASD in many studies. Thus, in terms of ecological validity, many researchers recommend the use of dynamic and time limited facial emotion stimuli in future research (Gepner and Féron 2009; Parish-Morris et al. 2013; Rahko et al. 2012; Uljarevic and Hamilton 2012). We will address this issue later in the “[Method](#)” section.

From the first study investigating explicit ToM in ASD (Baron-Cohen et al. 1985) to the latest meta-analyses concerning neural correlates of explicit ToM in ASD (Aboulafia-Brakha et al. 2011; Devine and Hughes 2014; Sugranyes et al. 2011), the deficient ability of inferring the mental states of others is a consistent observation in individuals with ASD. Nevertheless, the extent to which differences between ASD and NTD can be measured depends on the level of difficulty of the explicit ToM-tasks. While typically developed children

between the age of 3–4 years are able to solve verbally presented first-order FBT (Wellman et al. 2001; Wimmer and Perner 1983), most children with ASD do not pass this task until the age of 11 years (Happé 1995). In a test using a method of implicit learning, typically developed children aged 2 years are able to spontaneously anticipate others’ actions, while adults with ASD still fail to pass the test (Senju et al. 2010). Second-order FBT can be solved by children at the age of 6 and 7 years (Perner and Wimmer 1985). Advanced ToM-abilities, needed for complex social situations in which wrong behavior has to be represented by a cognitive and an empathic affective component, first appear in typically developed children between the age of 9 and 11 (Baron-Cohen et al. 1997).

The process of perception and categorization of social stimuli is an adaptive function, and in addition to facial emotional expression, it is based on the integration of several sources containing voice cues like prosody, body cues like gestures and posture, odor, time, space, and environment (Murphy 2012). The basic mechanism of this cognitive integration process of context effects was investigated by Rumelhart and McClelland (1981, 1982). Their interactive activation and competition model, tested by means of target letter perception in words, gives evidence to the assumption that context has a bearing on perception and that duration and timing of the context in which a target appears improves its perceivability. Different constructivist approaches consider contextual information an important framework that can change the categorization of the same facial expression, depending on the contextual background (Kim et al. 2003, 2004). The categorization of facial expression can be influenced by odor (Leppanen and Hietanen 2003) or emotional context (Righart and de Gelder 2008) and, furthermore, it has been shown that contextual factors have equal or greater influence on categorization of emotional information than the facial expression itself (Carroll and Russell 1996).

To assess the full range of demands on SC we compiled a test battery covering a broad range of well established and newly developed tasks that will put to the test the concept of implicit and explicit ToM with (1) facial emotion categorization (implicit), (2) first and second-order FBT (explicit), and (3) complex SC (tasks with complex social interaction in a contextual framework of a coherent story including implicit and explicit aspects of SC).

Compiling this detailed test battery, we investigated which differences arise between adolescents with high-functioning ASD and their typically developed peers. The aim was to determine if there is an overall deficit in SC, especially in ToM abilities of the participants with ASD as compared to NTD. Furthermore, we examined to what extent differences between implicit (concerning facial emotion categorization)

and explicit (concerning FB conditions) ToM measures correlate with autistic symptomatology and/or IQ.

Method

Participants

A total of 47 adolescents aged between 14 years 0 months and 17 years 11 months took part in the present study. Two participants had to be excluded from the original ASD group ($n=25$): One participant performed at chance level in all facial emotional recognition tasks conceivably based on prosopagnosia, and a second participant had to be excluded due to technical problems during the test session. Finally, 22 neurotypically developed participants (NTD) and 23 participants with ASD were included in the statistical analyses. ASD and control participants were recruited from project databases of the Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics of the Medical Center—University of Freiburg. The groups were matched by chronological age and nonverbal intelligence (see Table 1) as assessed by the CFT 20-R (Weiß 2008) that belongs to the family of culture fair intelligence test measuring general fluid intelligence (Cattell 1963). ASD diagnoses were established by means of instruments considered to be the gold standard, the Autism Diagnostic Observation Schedule (ADOS; German version: Rühl et al. 2004) and the Autism Diagnostic Interview-Revised (ADI-R; German version: Bölte et al. 2006a). To assess severity of symptoms of ASD concerning social responsiveness, we used the Social Responsiveness Scale (SRS; German version: Bölte and Poustka 2008). NTD had to fill out the CBCL/4-18 (Döpfner et al. 1994) a questionnaire for parents to assess behavioral and emotional problems of children, and its equivalent self-rating questionnaire YSR/11-18 (Döpfner et al. 1998) for exclusion of psychiatric symptoms. To classify empathizing and systemizing abilities based on the E-S-Theory of Baron-Cohen (2009), participants had to fill out the EQ and the SQ (German versions). Measuring alexithymia, assessing the inability to identify and describe one's own emotions and those of others, we used the Toronto Alexithymia Scale (TAS-26; German version; Kupfer et al. 2001). The study was conducted according to the principles expressed in the Declaration of Helsinki. Informed written consent was obtained before testing. The local ethics committee approved the study.

Tests

(1) Complex Tasks

MASC (Dziobek et al. 2006). The Movie for the Assessment of SC (MASC) is a sensitive video-based test to

Table 1 Sample's characteristics

	NTD ($n=22$)	ASD ($n=23$)	<i>F</i>	<i>p</i>	<i>g</i>
Gender	All male	All male			
Age	15.85 (0.97)	15.72 (1.25)	<1		
IQ	103.77 (11.09)	105.65 (11.47)	<1		
SRS-T-Total	42.23 (9.76)	74.87 (7.19)	164.15	<.0001	3.75
SRS-T-Awr	43.73 (10.40)	69.61 (7.95)	88.50	<.0001	2.76
SRS-T-Cog	44.91 (9.11)	70.22 (7.38)	105.35	<.0001	3.01
SRS-T-Com	45.09 (7.23)	79.65 (10.01)	174.96	<.0001	3.88
SRS-T-Mot	46.27 (9.91)	70.43 (10.47)	63.09	<.0001	2.33
SRS-T-RRB	49.77 (5.34)	71.78 (5.81)	174.83	<.0001	3.87
EQ	0.94 (0.22)	0.70 (0.29)	9.85	.003	0.92
SQ	0.71 (0.20)	0.68 (0.23)	<1		
TAS-26-Total	38.36 (6.73)	45.32 (7.63)	10.28 ^a	.003	0.95
TAS-26-DIF	10.82 (3.43)	14.64 (5.01)	8.70 ^a	.005	0.87
TAS-26-DDF	11.68 (3.50)	14.96 (3.51)	9.83	.003	0.92
TAS-26-EOT	15.86 (3.37)	15.65 (3.26)	<1		

NTD neurotypical development, ASD autism spectrum disorder, SRS social responsiveness scale, Awr social awareness, Cog social cognition, Com social communication, Mot social motivation, RRB restricted interests and repetitive behavior, EQ empathy quotient, SQ systemizing quotient, TAS-26 Toronto alexithymia scale, DIF difficulty identifying feelings, DDF difficulty describing feelings, EOT externally-oriented thinking, *g* Hedges's *g*

^a $n(\text{ASD})=22$, because 1 participant had too many missing values on this subscale

assess mindreading difficulties in complex social situations. Watching a 15 min movie about four individuals getting together for a dinner party, the video stops 51 times and the participants are required to answer several questions about the mental states of the different characters in a multiple choice paradigm. Main outcome is the accuracy rate for correctly given answers.

AToM (Schaller and Rauh 2011). The Animated-Theory-of-Mind (AToM) test is aimed at increasing ecological validity by using dynamic video-based stimuli close to everyday life (by means of the TV-animation series "Shaun the Sheep"). Within a non-verbal story that features social interactions, participants are asked about different aspects of SC including facial emotion expression, intentions, deception and first- and second-order false beliefs of protagonists (see Fig. 1). After every relevant scene the participants were asked questions about the facial expressions, thoughts, beliefs, and intentions of the protagonist(s) that are observable or inferable in a contextual and coherent story. Items concern FEC, FB, and comprehension of complex SC. The mean of these items serves as the AToM-Total score that integrates all SC-related answers. In addition,



Fig. 1 Scene from the Animated-Theory-of-Mind Test (AToM). The sheep deceive the farmer into abandon the swimming-pool, by letting him believe that a storm is brewing (printed with permission from Aardman Animations Limited. Shaun the Sheep (word mark) and the characters “Shaun the Sheep” (C) and TM Aardman Animations Limited. Licensed by WDR mediagroup GmbH)

participants are asked questions about the presence and number of peripheral objects (PO) outside the (social) gist of the scene. PO items are included to serve as a measure for basic aspects of story understanding of the participants beyond items concerning SC. On the one hand PO serves as an subscale of working memory, representing a measure to compare this aspect of executive functions (Miyake et al. 2000) between ASD and NTD. On the other hand we want to examine whether individuals with ASD show a basic understanding of the task respectively pay attention to the task at all.

In addition to the AtoM-Total score, the task allows for the assessment of six subscales: Facial Emotion Categorization (FEC=14 items; i.e. “What facial expression does the dog show when he recognizes the shark in the pool?”), first-order false belief (FB1O=16 items; i.e. “What does the farmer think is swimming in the pool?”), second-order false belief (FB2O=6 items; “What does Shaun think, what the farmer thinks, is swimming in the pool?”), comprehension questions concerning the causes for the emotional expression of the protagonists [Comprehension Facial Emotion Categorization (AToM-Comp-FEC)=13 items; i.e. “Why does he show this facial expression?”] and reality questions (AToM-Comp-RQ=4 items; i.e. “Who is actually swimming in the pool?”) concerning distinctions between appearance (beliefs of protagonists) and reality (perspective of the omniscient observer), and peripheral objects (AToM-PO=19 items; i.e. “How many objects have you seen standing on the side table next to the dog?”). The AToM-FEC subscale includes 14 FEC items. The AToM-FB subscale includes the items of FB1O and FB2O and thus comprises of 23 items. In this version of AToM, all answers are given verbally without a fixed response format and are audio-recorded. Analysis

was carried out by two independent blinded raters, assigning the verbally given answers using a prescribed list of possible answers for each question. Inter-rater reliability was assessed for (i) consistency as well as for (ii) absolute agreement using two-way random, single-measures ICC (McGraw and Wong 1996) for all AToM scales. According to the guidelines of Cicchetti (1994), the resulting ICC values were fair to excellent for consistency [AToM-Total: $ICC(C,I)=0.892$; AToM-FEC: $ICC(C,I)=0.808$; AToM-FB: $ICC(C,I)=0.664$; AToM-Comp-FEC: $ICC(C,I)=0.894$; AToM-Comp-RQ: $ICC(C,I)=0.538$; AToM-PO: $ICC(C,I)=0.832$] as well as for absolute agreement [AToM-Total: $ICC(A,I)=0.788$; AToM-FEC: $ICC(A,I)=0.747$; AToM-FB: $ICC(A,I)=0.606$; AToM-Comp-FEC: $ICC(A,I)=0.828$; AToM-Comp-RQ: $ICC(A,I)=0.542$; AToM-PO: $ICC(A,I)=0.802$].

In cases of ambiguity a third independent and blinded rater decided if the answer should be counted as right or wrong. Main outcome is the accuracy rate for correctly given answers.

(2) Implicit Tasks

Dynamic Emotion Categorization Test—DECT (Rauh and Schaller 2009). In order to prove the usefulness of dynamic facial stimuli of emotion, we combined videos of natural actors with animated artificial actors displaying facial expressions of basic emotions at a stated level of intensity. Four characters (two natural/two virtual) express six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) with three levels of intensity (weak, medium, strong). In a pilot study, a certified coder of the facial action coding system (Ekman and Friesen 1976) had screened the 72 items for consistency with the FACS. One female and one male virtual character were designed and specified with the Agent Framework (Helzle et al. 2004). Each video clip or animation lasts for 6 s, wherein for the first 3 s the corresponding face is neutral and then starts to progress into the emotional expression. The Dynamic Emotion Categorization Test (DECT) starts with seven practice trials followed by the 72 video clips or animations. Facial emotion type and intensity level were distributed in a pseudo-randomized manner with the restriction that in one block of trials all six basic emotions were presented exactly once, and each intensity level (weak, medium, strong) exactly twice. Responses were given verbally and the experimenter enters the verbally given answer per button press. For every item six basic emotions were given (anger, disgust, fear, happiness, sadness, surprise). Main outcome is the accuracy rate for correctly categorized emotions.

Faces-Test (Baron-Cohen et al. 1997). In the Faces-Test b/w photographs of an actress expressing seven “basic” emotions (afraid, angry, disgusted, distressed, happy, sad,

and surprised), and nine “complex” mental states (admiring, arrogant, bored, flirting, guilty, interested, quizzical, scheming, and thoughtful) are used. Each photo is accompanied by a target adjective and an incorrect adjective (e.g., “admiring” versus “arrogant”), describing the depicted mental state of the actress. The original paper-pencil version was implemented in Presentation® (Neurobehavioral Systems Inc., Berkeley, USA) and modified with regard to time limitation. The pictures are presented for four duration conditions (150, 300, 450 ms, self-paced) in pseudo-randomized order, and the participant has to make his decision in a forced choice procedure between the two alternatives (target state and foil alternative state) by pressing one of two keys on the numeric keypad. Main outcome is the accuracy rate for correctly categorized emotions.

RMET (Baron-Cohen et al. 2001). Similar to the Faces-Test, the computerized child version of the Reading-the-Mind-in-the-Eyes Test (RMET) uses 36 b/w photographs. Instead of whole faces, the RMET only shows the eye-region of a face expressing an emotion depicting a distinguishable mental state. Every picture is accompanied by four words, describing mental states (e.g., “bored”, “interested”, “relaxed”, “thoughtful”). The original paper-pencil version was also implemented in Presentation® (Neurobehavioral Systems Inc., Berkeley, USA). The main outcome is the proportion of target selections (versus non-target selections). Responses were recorded by pressing one of four keys on the numeric keypad.

FEEL (*Facially Expressed Emotion Labeling*; Kessler et al. 2002). The FEEL test (Version 3.2) is a computer-based test measuring perception of facially expressed emotions. All six basic emotions developed by Matsumoto and Ekman (1988; JACFEE, anger, disgust, fear, happiness, sadness, and surprise) are presented on a computer screen for 300 ms. For the first 1500 ms the stimulus character depicts a neutral face, and then switches into one of the six basic emotions for 300 ms. After image presentation, the participants choose the corresponding emotion via mouse click within a time interval of 10 s. Main outcome is the accuracy rate for correctly categorized emotions.

FEFA (*Frankfurt Test and Training of Facial Affect Recognition*; Bölte et al. 2006b). A computer-based emotion recognition test consisting of 50 b/w photographs of faces depicting all six basic emotions and a neutral expression. Each picture is shown on the screen and the participant is asked to select one out of six emotions or the description “neutral” by mouse click. Main outcome is the accuracy rate for correctly categorized emotions.

(3) Explicit Tasks

Classical FBT (sensu Wimmer and Perner 1983). In order to test the comprehension of FB of others, three different

stories of first ($n=1$) and second order ($n=2$) FB were presented. The first-order FB story is a slightly adapted version of the classical “Sally & Anne-Task” by Baron-Cohen et al. (1985). The two second-order FB stories are a narrowly modified adaptation of the “ice cream van-story” by Perner and Wimmer (1985) and the so called “Waldgeschichte” (“wood-story-task”) by Wichmann (1996). In all three stories the experimenter asks a false belief question (“Where will person A look for object X/person B?”), an explanation question (“Why?”), a control question (“Where is object X/person B really?”), and a knowledge question (“Does person A know that person B has hidden object X”). In case of second-order FB stories there are additional questions: a question about the state of not-knowing (“Does person A know that person B knows?”) and a second-order belief question (“What does person A believe what person B believes?”). This procedure assesses whether subjects have an explicit representation of another person’s FBs (Wimmer and Perner 1983). Main outcome is the accuracy rate for correctly given answers.

Procedure

Individuals who fulfilled inclusion criteria were invited to participate in the study. Subsequently, they had to complete the above mentioned compilation of surveys (all: SRS, EQ, SQ, TAS-26; NTD only: CBCL/4-18, YSR/11-18). Adolescents were assessed within two sessions, separated by a maximum interval of 14 days. All participants completed the test battery individually in a quiet room. In the first session with a total duration of approximately 2 h, the paper and pencil short form of CFT 20-R was conducted within a time frame of about 30 min, followed by the computerized DECT, RMET, Faces-Test, FEEL, and FEFA. In the second session, the participants had to complete the second part of the SC test battery including MASC, AToM and classical FBT within approximately one and a half hours.

Measures

Computerized stimuli were presented on a 17 inch monitor connected to a PC running Presentation® software (Version 15.1, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) on a Microsoft Windows XP operating system. Group differences were analyzed by means of ANOVA. In case of violations of distributional assumptions of parametric tests, non-parametric alternatives are applied. Checking assumptions of normality are conducted with Kolmogorov–Smirnov tests. In the case of accuracy rates, the well-established arcsine transformation (aka “arcsine square root transformation” or “angular transformation”) of proportional values was applied in order to see whether that remedies the normality violations (e.g., Hair et al. 2014). If

so, ANOVAs were applied to the arcsine-transformed data; if not, non-parametric Wilcoxon–Mann–Whitney tests were computed. For the purpose of distinctions between ASD and NTD we computed analyses of all implicit, explicit and complex measures of the SC test battery. In order to investigate convergent and concurrent relations of implicit, explicit and complex tasks we computed Pearson correlations of all accuracy outcomes. Furthermore, we examined if there are any developmental effects of age on the performance in SC considering the above mentioned measures (implicit, explicit, complex tasks). For this we ran Pearson correlations for the total sample and each group separately. Effect sizes for group differences are reported in terms of standardized mean differences (SMD): Unbiased Hedges's g , rather than Cohen's d , is used as point estimator of effect sizes (Borenstein et al. 2009), because the former enables the computation of the 95% CI that go also into the forest plot of the systematic review of results. All statistical analyses are performed with SAS software, Version 9.4 (SAS Institute Inc., Cary, NC, USA). For all statistical analyses, a significance level of $\alpha = .05$ is adopted.

Results

Sample's Characteristics

In Table 1, the sample's characteristics are presented. There were no significant group differences for age and IQ [both $F(1, 43) < 1$]. Concerning autistic symptomatology, the ASD group showed significantly higher T-scores for the SRS and its subscales—with large effect sizes (all $g_s \geq 2.33$). In addition, alexithymia scores as assessed by the TAS-26 were also significantly higher for the ASD group, except for the TAS-26-EOT (= Externally-Oriented Thinking) subscale. Concerning empathizing and systemizing, results were mixed: ASD adolescents reported significantly lower empathy (lower EQ scores), whereas no significant group difference was found concerning degree of systemizing as assessed by SQ.

Group Differences in Measures of SC

In the following, the results of the behavioral tests concerning social cognition (complex tests, implicit tests, and explicit tests) will be reported. In Table 2, descriptive (means and standard deviations) and the inferential statistics are summarized.

Complex Test—AToM

Considering the compound AToM-Total Score (assessing complex social cognition), the ASD group ($M=0.852$,

$SD=0.117$) scored significantly lower than the NTD group ($M=0.936$, $SD=0.057$) [$F(1,43)=9.27$, $p=.004$, $g=-0.89$]. Since the assumption of normal distribution was violated before and after arcsine transformation of the AToM-FEC subscale, non-parametric tests were applied. The Wilcoxon–Mann–Whitney test yielded a significant difference between both groups (NTD: $M=0.912$, $SD=0.085$; ASD: $M=0.826$, $SD=0.150$; $W_s=593.50$, $z=2.024$, $p=.043$). For the AToM-FB subscale the assumption of normal distribution was also violated before and after arcsine transformation. The Wilcoxon–Mann–Whitney test revealed a significant difference between both groups (NTD: $M=0.981$, $SD=0.030$; ASD: $M=0.934$, $SD=0.068$; $W_s=608.00$, $z=2.449$, $p=.014$).

Additional to the items affecting SC, we examined potential differences between groups on PO items in the AToM without obtaining any significant effects [NTD: $M=0.512$, $SD=0.123$; ASD: $M=0.487$, $SD=0.121$; $F(1,43) < 1$].

Complex Test—MASC

For the MASC (assessing complex social cognition), the total score in the ASD group deviated significantly from the normal distribution. Therefore, an arcsine transformation was applied before parametric testing. ASD adolescents were more inaccurate when responding to mental state questions [$F(1,43)=6.36$, $p=.015$, $g=-0.78$].

Implicit Test—DECT

One-way ANOVA with the main outcome (overall accuracy) revealed significant deficits in (dynamic) facial emotion categorization for the ASD group [$F(1,43)=5.66$, $p=.022$, $g=-0.70$].

Implicit Test—Faces-Test

For the Faces-Test (assessing FEC), the total accuracy score in each group deviated significantly from the normal distribution. Therefore, an arcsine transformation was applied before parametric testing. Upon computing a one-way ANOVA for the Faces-Test, neither significant differences between groups for the arcsine-transformed overall accuracy rate [$F(1,43)=2.44$, $p=.126$], nor for the time limited sub-indices (150 ms, 300 ms, 450 ms, self-paced) were found [all $F_s < 1$, except for the 450 ms condition: $F(1,43)=3.18$, $p=.082$].

Implicit Test—RMET

Considerable differences in facial emotion categorization were found between ASD and NTD in the RMET

Table 2 Mean accuracy rates for implicit, explicit, and compound measures of social cognition

	NTD (<i>n</i> =22)	ASD (<i>n</i> =23)	Test statistic	<i>p</i>	<i>g</i>
Implicit					
DECT	0.649 (0.044)	0.615 (0.051)	$F(1,43)=5.66$.022	-0.70
Faces-Test ^a	0.845 (0.086)	0.796 (0.121)	$F(1,43)=2.44$.126	-0.47
RMET	0.714 (0.090)	0.585 (0.124)	$F(1,43)=15.84$	<.001	-1.17
FEEL	0.827 (0.086)	0.793 (0.112)	$F(1,43)=1.28$.264	-0.33
FEFA	0.860 (0.053)	0.805 (0.073)	$F(1,43)=8.27$.006	-0.84
Explicit					
Classical FB ^b	0.944 (0.089)	0.894 (0.144)	$W_s=569.50$.114	-0.41 ^c
Compound					
MASC ^a	0.710 (0.084)	0.598 (0.180)	$F(1,43)=6.36$.015	-0.78
AToM-Total	0.936 (0.057)	0.852 (0.117)	$F(1,43)=9.27$.004	-0.89
AToM-FEC ^b	0.912 (0.085)	0.826 (0.150)	$W_s=593.50$.043	-0.69 ^c
AToM-FB ^b	0.981 (0.030)	0.934 (0.068)	$W_s=608.00$.014	-0.86 ^c
AToM-Comp-FEC ^b	0.916 (0.089)	0.796 (0.213)	$W_s=602.00$.026	-0.72 ^c
AToM-Comp-RQ ^b	0.977 (0.074)	0.989 (0.052)	$W_s=494.00$.546	0.18 ^c
AToM-PO	0.512 (0.123)	0.487 (0.121)	$F(1,43)<1$		0.20

NTD neurotypical development, ASD autism spectrum disorder; DECT dynamic emotion categorization test, RMET reading-the-mind-in-the-eyes test, FEEL facially expressed emotion labeling test, FEFA Frankfurt test and training of facial affect recognition, FB false belief, MASC movie for the assessment of social cognition, AToM animated theory-of-mind, AToM-FEC AToM subscale facial emotion categorization, AToM-FB AToM subscale false belief, AToM-Comp-FEC AToM subscale comprehension of facial emotion categorization, AToM-Comp-RQ AToM subscale comprehension of reality questions, AToM-PO AToM subscale peripheral objects, *g* Hedges's *g*

^aANOVA is based on arcsine-transformed scores [$X_{\text{new}} = 2 \arcsin(\sqrt{X_{\text{old}}})$]

^bWilcoxon–Mann–Whitney-test

^cValues of *g* may be biased because of violations of the normality assumption

and yielded significant results in the one-way ANOVA [$F(1,43)=15.84$, $p=.0003$, $g=-1.17$]. Performance of the ASD group was much lower than in the NTD group (see also Table 2).

Implicit Test—FEEL

No significant differences were found for the FEEL test (that also assesses FEC) comparing both groups for accuracy in a one-way ANOVA [$F(1,43)=1.28$, $p=.264$]. Separate analyses for accuracy rates of the six basic emotions also yielded no significant group differences.

Implicit Test—FEFA

For the FEFA, the one-way ANOVA revealed distinct deficits in facial emotion categorization for the ASD group [$F(1,43)=8.27$, $p=.006$, $g=-0.84$]. Moreover, regarding the accuracy rates for the 6 basic emotions, differences were significant for anger [$F(1,43)=6.48$, $p=.015$], and for fear [$F(1,43)=6.11$, $p=.018$], but not for disgust [$F(1,43)=1.22$, $p=.276$], happiness [$F(1,43)<1$], sadness [$F(1,43)=2.09$, $p=.156$], and surprise [$F(1,43)<1$].

Explicit Test—Classical FBT

Since the assumption of normal distribution was violated before and after arcsine transformation, non-parametric tests were applied. Wilcoxon–Mann–Whitney tests yielded no significant differences between both groups [neither in FB1O ($W_s=517.00$, $z=0.551$, $p=.581$) nor FB2O ($W_s=575.50$, $z=1.748$, $p=.081$)]. The FB-Total score (NTD: $M=0.944$, $SD=0.089$; ASD: $M=0.894$, $SD=0.144$) was also not significant ($W_s=569.50$, $z=1.579$, $p=.114$), indicating that both groups at least do not differ widely in their ability to detect false beliefs in stories presented in verbal format.

Systematic Review of Results

In Fig. 2, group differences for all measures are displayed in a forest plot to give a concise systematic review of obtained results. All measures were scaled so that results falling to the right of the line-of-no-effect (zero) indicate worse performance, more symptomatology, more systemizing or less empathizing than the NTD group. As can be seen, nearly all measures, except for SQ and TAS-26-EOT, are descriptively heading for the hypothesized direction.

Significant differences are obtained for (a) the implicit tests DECT, RMET, and FEFA, (b) for the complex tests MASC and AToM, and (c) for the AToM subscales AToM-FEC, AToM-FB, and AToM-Comp-FEC. Within the group of questionnaires, the SRS together with its five subscales yielded the largest effects (all $gs > 2.33$). In addition, EQ as well as the TAS-26 together with the two subscales TAS-26-DIF and TAS-26-DDF are significant. In contrast, the 95% CIs of TAS-26-EOT and SQ cover the line-of-no-effect.

Correlations Between Measures of SC

In Table 3, the correlations of the implicit facial emotion categorization tests together with the explicit test and the tests of complex social cognition (MASC and AToM scales) are displayed (skipping the social cognition unrelated AToM subscales AToM-Comp-RQ and AToM-PO): Concerning FEC, there are roughly three groupings of noteworthy intercorrelations: (1) All pairwise correlations of DECT, Faces-Test and RMET are significant, (2) FEEL and FEFA also correlate significantly. (3) In contrast, the AToM-FEC subscale does not correlate with any of the other implicit tests. The only exception to this grouping is the significant correlation of RMET with FEFA ($r = .380, p = .010$).

Concerning the explicit test, the classical FBT correlates significantly with DECT ($r = .372, p = .012$) and RMET ($r = .365, p = .014$), but also with the complex measures

of MASC, AToM-Total und AToM-FB ($r = .443, p = .002; r = .341, p = .022; r = .325, p = .029$, respectively). For the total scores of the two complex measures of SC there is a significant correlation of MASC with AToM-Total ($r = .509, p < .001$), that is considered as a large effect according to common statistical conventions (Cohen 1988). In addition, all scales of the compound tests correlate significantly with each other as can be seen in Table 3. However, the intercorrelations should be interpreted with caution, because the correlations are based on a rather small sample. Therefore, we also refrained from running principal component analyses (PCA), because results hinges from honest correlations, and sample size is definitely below recommended rules of thumb: Hair et al. (2014) for example state: “The minimum absolute sample size should be 50 observations” (p. 100).

Effects of Age on SC

In order to consider tentative influences of age on performance in SC, we computed Pearson correlations for the total sample and each group separately, for all measures of SC with age.

With regard to the five context-free FEC tests, there were two tests (RMET, Faces-Test) for which the total score correlated significantly with age [RMET: $r = .304, p = .043$; Faces-Test (arcsine-transformed): $r = .326, p = .029$]. In both cases, the correlation within the NTD group was not significant [RMET: $r = .074, p = .744$;

Fig. 2 Forest plot of standardized mean differences for all instruments (total scores and subscales) applied in the study

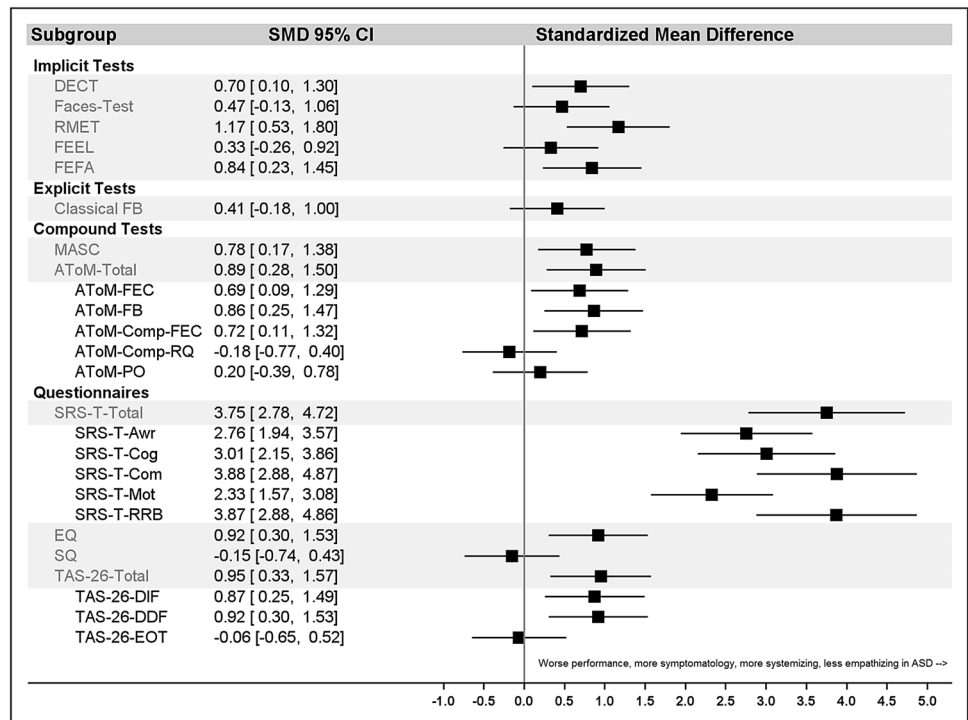


Table 3 Summary of intercorrelations, means, and standard deviations of accuracy rates for implicit, explicit, and compound measures of social cognition

Measure	1	2 ^a	3	4	5	6 ^a	7 ^a	8	9 ^a	10 ^a	11 ^a	<i>M</i>	<i>SD</i>
Implicit													
1. DECT	–	.467**	.592***	.209	.132	.372*	.499***	.257	.157	.184	.183	0.632	0.050
2. Faces-Test ^a		–	.421**	.281	.274	.248	.377*	.112	–.123	.010	.235	2.295	0.276
3. RMET			–	.188	.380*	.365*	.513***	.310*	.143	.352*	.299*	0.648	0.126
4. FEEL				–	.592***	.073	.136	–.151	–.233	–.039	–.041	0.810	0.101
5. FEFA					–	.077	.374*	.183	.125	.236	.228	0.832	0.069
Explicit													
6. Classical FB ^a						–	.443**	.341*	.214	.325*	.284	2.743	0.459
Compound													
7. MASC ^a							–	.509***	.371*	.520***	.323*	1.893	0.331
8. AToM-Total								–	.811***	.577***	.885***	0.893	0.101
9. AToM-FEC ^a									–	.345*	.665***	2.517	0.445
10. AToM-FB ^a										–	.315*	2.848	0.311
11. AToM-Comp-FEC ^a											–	2.502	0.544

DECT dynamic emotion categorization test, RMET reading-the-mind-in-the-eyes test, FEEL facially expressed emotion labeling test, FEFA Frankfurt test and training of facial affect recognition, FB false belief, MASC movie for the assessment of social cognition, AToM animated theory-of-mind, AToM-FEC AToM subscale facial emotion categorization, AToM-FB AToM subscale false belief

* $p < .05$

** $p < .01$

*** $p < .001$

^aArcsine-transformed scores [$X_{\text{new}} = 2 \arcsin(\sqrt{X_{\text{old}}})$]

Faces-Test (arcsine-transformed): $r = .231$, $p = .301$], whereas for the ASD group both correlations showed (tendency for) a significant correlation [RMET: $r = .450$, $p = .031$; Faces-Test (arcsine-transformed): $r = .375$, $p = .078$]. In addition, all correlation coefficients in the other context-free FEC tests were descriptively greater in the ASD group than in the NTD group.

For the explicit task [Classical FB-Total Score (arcsine-transformed)], no significant correlation with age was found ($r = -.175$, $p = .250$). Within groups, there was an unexpected significant negative correlation in the NTD group ($r = -.444$, $p = .038$), mostly due to the oldest NTD participant who performed rather poorly in the classical FBT. After removing this most influential observation the significant correlation disappears ($r = -.300$, $p = .186$). For the ASD group, no significant correlation was found ($r = -0.044$, $p = .842$).

For the complex scores of SC (MASC, AToM), the pattern of results was mixed: For the MASC (arcsine-transformed Total score) there was a significant correlation with age for the whole sample ($r = .362$, $p = .015$). Within groups, no significant correlation was obtained for the NTD group ($r = .193$, $p = .391$); for the ASD group, however, there was a significant correlation with age ($r = .438$, $p = .037$). In Fig. 3, the differential developmental trajectories are depicted in a scatterplot.

For the AToM (Total score), neither a significant correlation with age was obtained for the total sample ($r = -.006$, $p = .967$); nor for the two subsamples of ASD ($r = -.072$, $p = .745$); and NTD ($r = .059$, $p = .795$).

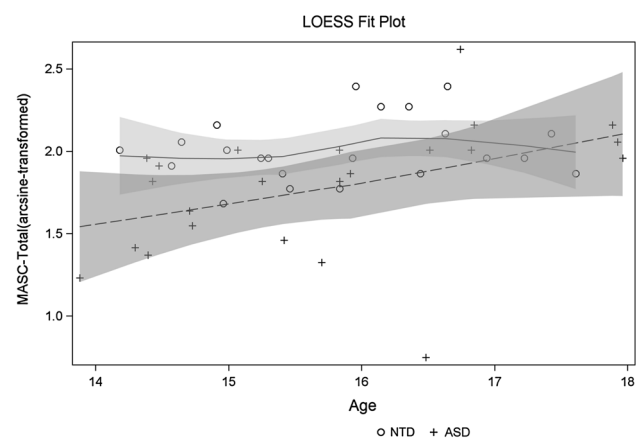


Fig. 3 Scatterplot of MASC-Total scores (arcsine-transformed) and age of participants, together with LOESS curve and 95% CI. For the ASD group, a linear increase of MASC performance with age can be noted, whereas for the NTD group a plateau seemed to be reached already at the age of 14

Discussion

The aim of the present study was to investigate SC in order to assess performance differences between adolescents with ASD and neurotypical controls on the basis of a multifaceted test battery. Comprised of eight behavioral tasks, the ToM test battery addresses implicit (FEC), explicit (false belief), and complex aspects of SC.

We found significant differences between ASD and NTD in three of five tasks concerning iToM respectively the perception and categorization of facial emotion without any contextual factors. In contrast, there were no significant eToM differences in the performance measures of classical FBT (first and second-order). In the case of complex social cognition tasks we found significant differences in the newly developed AToM test and in the well established MASC.

Additionally, in the area of FEC we found significant correlations between DECT, Faces-Test and RMET, between RMET and FEFA and between FEFA and FEEL (see Table 3). The influence of age on SC, especially on implicit ToM and complex SC abilities was descriptively higher for ASD in all tasks and yielded significance in correlations with RMET and Faces-Test (FEC) and MASC (complex SC).

SC as the dynamic constructive and conceptual processing of the social environment arises as a result of multimodal perception. Obviously, alert human individuals are predominantly exposed to a more or less permanent flow of social information in everyday life. The incoming perceptions must be handled efficiently within a limited timeframe and without having full access to all necessary information (Hertwig and Hoffrage 2013). In order to efficiently and successfully develop personal expectations that correspond to the expectations of the social environment, the SC system depends on the implicit processing of the inrushing stimuli. Combining paradigms measuring implicit as well as explicit and complex responses in the test battery, our results suggest that—as adolescents with ASD transition into adulthood—they still display an overall deficit in social cognition.

Concerning general task performance, the comparison of accuracy between ToM-items and distractive PO-items in AToM suggest that participants with ASD do have deficits in SC regardless of their general effort in the task, because their performance is as good as the NTD group in the PO-items.

These findings are reflected exemplarily in the newly developed AToM test, which shows significant differences of performance between ASD and NTD in FEC-items as well as in FB-items in complex social conditions. The compound structure of the test provides conclusively more processable information than an emotional expression of a face

in a noncontextual task can yield. Thus, we suggest that embedding implicit and explicit aspects of social cognition in a continuous/sequential story poses a higher demand for ASD than explicit false belief stories or static and exaggerated displays of facial emotion.

Implicit

In accordance with previous research we can mainly find significant differences in the performance between autistic and neurotypical individuals concerning facial emotion categorization (Harms et al. 2010; Lozier et al. 2014; Uljarevic and Hamilton 2012). On closer examination of implicit tasks it is remarkable that there is a wide range of results consisting of small differences on the descriptive level (FEEL, Faces-Test) as well as significant differences with large effect sizes (RMET, FEFA, DECT). Especially the RMET yielded considerable performance disparities between NTD and ASD. It seems plausible that these differences can be ascribed to the piecemeal reduction of relevant mimic information and to the lack of any contextual information. Since the ability to relate the fragments of facial expressions of mental states to mental concepts is described as unconscious, rapid, and automatic (Todorov 2012), the distinct deficits of the ASD group can be regarded as strong evidence for impairments concerning implicit facial emotion processing.

Looking at it the other way round, the addition of contextual information to a reduced social stimulus then would facilitate the categorization of the mental state in the therefore provided lexicon. This could be an explanation for the minor differences between NTD and ASD in the Faces-Test and in the FEEL where facial stimuli are presented in an entire, exaggerated and static form. And as Baron-Cohen et al. (2001) stated, the results of ASD in basic emotion recognition tests may be camouflaged by learning compensatory strategies and lead to the fallacious assumption that individuals with high-functioning autism spectrum conditions show no deficits in facial emotion categorization.

Besides, the presentation of fragmentary faces in the RMET, some of the implicit tests in the SC test battery are characterized by the use of dynamic stimuli and virtual characters (DECT) as well as time-limited presentation of stimuli (Faces-Test, FEEL). Although NTD performed significantly better than ASD, the accuracy rates in the DECT are distinctly lower in both samples. Dynamic virtual characters especially in a context-free setting seem to be more challenging to categorize compared to static photographs of exaggerated facial expressions. Looking at previous research it can be assumed that dynamic stimuli reveal subtle emotion categorization deficits (Evers et al. 2015; Kessels et al. 2010; Sato et al. 2013; Uono et al. 2009). Moreover, imaging studies using dynamic stimuli could

show, that individuals with ASD compared to NTD generate different activity patterns in the human face processing system including the amygdala, fusiform gyrus and posterior superior temporal sulcus region (Pelphrey et al. 2007). Nonetheless, it is notable that previous research comparing static versus dynamic stimuli has yielded mixed results in NTD: (i) they show performance loss (Brosnan et al. 2015) and (ii) there is no significant difference to ASD participants (Bekele et al. 2014). Furthermore, it is conceivable that especially dynamic basic emotions with weak or medium intensities are very difficult to categorize for both NTD individuals and participants with ASD without having contextual factors like voice, gesture, posture, and situation. There is no doubt that contextual information can be helpful in situations of ambiguity, but in case of absent contextual information about emotional states, one has to revert to solely facial information. That means the smaller the amount of decisive contextual information provided, the more important it is to be able to discriminate between minimal differences.

While the FEFA, using static, exaggerated stimuli presented without time limit, shows a significant difference between ASD and NTD, especially those tests using time restrictions (FEEL, Faces-Test) did not result in significant group differences. Nonetheless, our findings are similar to those of another study presenting facial emotion stimuli with limiting exposure time that failed to elicit differences between ASD and NTD (Tracy et al. 2010). This may be, at least in part, attributable to the fact that the durations of the stimulus presentations were too long to establish differences between ASD and NTD, both in the study of Tracy et al. (2010) with stimulus exposure times of 1500 ms and in our study with varying exposure times of 150, 300, and 450 ms. In contrast, the use of markedly shorter stimulus exposure time (15 and 30 ms) has led to significant differences in emotion categorization between ASD and NTD (Clark et al. 2008).

A further explanation for the missing significances in both implicit tests can be seen in the visual stimulus presentation using static depictions and exaggerated facially expressed emotions of insufficient ecological validity. Everyday social interaction depicts a wide range of rapidly displayed, subtle and nuanced emotional expressions, while most studies investigating social cognition in ASD are restricted to prototypical stimulus material. This methodological point of criticism can be found in several publications with increasing frequency (Clark et al. 2008; Rump et al. 2009; Sasson et al. 2015; Tracy et al. 2010; Uljarevic and Hamilton 2012). In contrast to the results of FEEL and Faces-Test, the FEFA as a test procedure using static, exaggerated stimuli presented without time limit, shows a significant difference between ASD and NTD, indicating that even tests with a lower ecological validity can elicit deficits

in emotion categorization in ASD. Furthermore it can be stated, that tasks concerning FEC show only reduced inter-correlations (see Table 3), thus the consistency of the construct of facial emotion categorization strongly depends on the structure and features of the tasks. This applies especially to the case of complex tasks, where we can see that the FEC-items of the AToM show no significant correlations with any of the implicit tasks, underlining that FEC embedded in a complex social event differs distinctly from FEC without any contextual information.

Explicit

Based on recent studies, adolescents and adults with high-functioning ASD apparently show typical performance in FBT as long as they are explicitly instructed (Baron-Cohen et al. 1999; Frith and Happé 1994; Senju 2013). These findings can be reconciled with the results of the classical false belief paradigm we used to test explicit ToM abilities, indicating no significant differences between ASD and NTD. Showing no impairments in the case of eToM, compensation seems to be possible somewhere in the age range from 14 to 18 years, while there are still substantial impairments in iToM at least until the age of 18.

If we take a closer look at the process of comparative perspective taking as is customary in FBT, knowledge about the wrong or missing information of the protagonist results primarily from the distinction between the information that the protagonist can have and the information the participant as observer of the whole scene has. In none of the items presented in the explicit FB stories the participants have to process bodily expressions like facial emotion expression, gesture, posture, or other decisive contextual social information. The only relevant aspect for a correct answer is the correct conclusion that the presence or absence of a protagonist is decisive, and that the protagonist has or has not perceived the same information as an observer of the entire scene. Referring to causal learning models, we suggest that participants solving a FBT can attend to causal properties. Causal representations provide relations and processes with a mechanistic basis for the correlations between protagonists, time and space. Following Johnson-Laird (1981), a participant using causal properties given in a simple explicit FB story can create a model of correlations between subjects and their agency, objects, time and space that interprets the intentions and beliefs of the protagonist. Causal learning models are basic elements for the development of social schemata. While NTD are able to build up a rich variety of social schemata continuously growing into modules of folk psychology, our data emphasizes that adolescents with ASD show at least a different development of social schemata leading to different interpretations of complex social scenes. However, the

used FB stories are of such primitive structure without any description of emotional expressions or social complexity, that they are solvable without drawing on fully shaped social schemata, instead using simple inferences based on explicitly given premises. Taking a look at the study of Jolliffe and Baron-Cohen (1999) using the “strange stories-task” (Happé 1994) it is ascertainable that one of the differences between classical FB stories and the “strange stories” is the use of more naturalistic and complex stories (Happé 1994) containing social scripts, social schemata, metaphorical expressions, sarcasm, and jokes. In the study of Jolliffe and Baron-Cohen (1999) the comparison between a causal reasoning task and the “strange stories-test” suggests, that even when individuals with ASD show normal performance in causal reasoning, they show deficits in the mentalizing task. In relation to the basic structure of the classical FBT it is assumable that the deficits arise primarily from the difficulty to understand the social scripts and schemata. Thus, individuals with ASD fail in everyday social interactions, because they are exposed to a continual stream of different ToM challenges. Confronted with this vast quantity of nonverbal and contextual cues, and incapable of processing them automatically, the congestion leads to a bottleneck of information processing of social stimuli. In contrast, NTD individuals are able to process social information mostly implicitly and procedurally having recourse to fully fledged social schemata.

Complex

With regard to the AToM test including both implicit and explicit stimuli in a contextual framework of sounds, music, vocalizations, course of action, spatial and environmental cues, it is remarkable that participants with ASD performed significantly worse in both ToM conditions than their controls. We attribute these differences in performance on the high demands arising from a complex task. It contains child-oriented, social-interactive but non-verbal stories with several interacting protagonists and their network of relationships, being combined with a catalogue of questions putting high demands on executive functions, working memory processes, attention in areas of facial emotion categorization, understanding of false beliefs as well as remembering objects and events outside the social gist. Neurotypically developed participants may have the ability to refer to facial emotions, gesture and posture in the context of a story without achieving high cognitive load. This cognitive effortlessness can be associated with the familiarity of a situation. A familiar course of an event is easier to capture and interpret than an entirely unknown situation. NTD are able to identify key properties and relations between objects and/or subjects. This ability results from the use of schemata and scripts (Schank and Abelson 1977). Besides

the above mentioned causal-learning-model, we especially use these schemata to organize commonly experienced social situations (Christensen and Michael 2016). The use of schemata and scripts is implicit and makes it possible to facilitate a complex social situation, by establishing a framework that integrates informations and emphasizes key relationships (Hoernig et al. 1993; Kintsch and van Dijk 1978). Especially social cues can be used by NTD for associative relations of stored scenarios, becoming aware in a fast and automatic way, whereas individuals with ASD have no or insufficient recourse to such prototypical mental schemata, especially if they are induced by emotional or social stimuli. These deliberations are consistent with a previous study comparing scripts for social routines between high-functioning ASD with NTD (Trillingsgaard 1999).

Similar to the AToM, the setup and structure of the MASC shows a complex framework of social interactions, embedded in a self-contained story including a variety of interaction patterns that have to be categorized and answered in a multiple choice task. However, in the MASC another component is important for the correct categorization of social interactions between protagonists. The plot of this task highlights social categories of friendship, personal relationships and their initiation in peer groups, thus the comprehension respectively the implicit and explicit knowledge of these social schemata and categories, in other words, the stereotypical roles of social groups (and their interactions), are decisive for constructing meaningful representations of others (Bodenhausen et al. 2012). From an evolutionary perspective, we have to regiment the consort with humans as a risky endeavor (Heatherston et al. 2004). For a successful course of this adventurous challenge, it is necessary that we are capable of categorizing the behavior of the counterpart correctly. First and foremost the basis for this is that long before verbal communication, we capture information from the bodily expression in the form of gesture and facial expression. In an interplay with the verbal information emerges a pattern of congruent and incongruent (e.g. irony) perceptions (Murphy 2012).

Categorizing these social perceptions and then again attributing them to larger clusters and social schemata, is not thought to be innate but a process of learning (Heatherston et al. 2004) starting from the very first day and being encouraged by parents, family, peers, milieu, and media. It is conceivable that there is a direct relation to iToM abilities and the understanding of such social schemata as friendship and personal relationship. Following the model of social development suggested by Soto-Icaza et al. (2015), the temporal perspective of biological and behavioral specialization according to social cognition is referable to a progressive development in social abilities dependent on the interaction between the social brain and the surrounding environment. While NTD have direct access to implicit,

automatic knowledge and scripts, it is assumable that individuals with ASD are lacking social concepts and schemata. The simultaneity of relevant and irrelevant events in the stories of AToM and MASC constrains individuals with ASD to manage the processing of the intruding information in an explicit, and conscious manner under high cognitive load. This may lead to an overextension in real social situations and deficits in accuracy in complex tasks.

Age Correlations

As indicated, age can be an important factor for accuracy in SC tasks. A closer look at correlations reveals significance for MASC, Faces-Test and RMET. Specifically the correlation between MASC as a complex task (see Fig. 3) and age as well as facial emotion categorization (RMET) and age can be found in the ASD group. This finding may imply that crucial and decisive steps in the development of facial emotion categorization and the development of social schemata are still incomplete in the ASD group, while neurotypically developed adolescents show a flatter learning curve.

Referring to the meta-analysis of Uljarevic and Hamilton (2012), impairments of FEC in ASD across age can be assumed. With respect to the age range between childhood and adolescence, many studies reveal distinct FEC deficits, while others cannot find significant differences in that age group (for a review, Harms et al. 2010). Neuroimaging and electrophysiological data suggest that face processing in adults can be traced back to earliest activation of brain areas specific for face detection mechanisms within the first months after birth (Grossmann et al. 2007, 2008).

The significant correlation between age and accuracy in the MASC is also solely attributable to the ASD group. Thus the developmental deficit can be associated with impairments in facial emotion categorization, leading to deficient answers in those items concerning facial emotion expression. On the other hand it might be possible that the developmental deficits of ASD show their influence when concurrent processing of various different stimuli of social and non-social kind needs recourse to elaborated social schemata.

Conclusions

In this study we applied several tasks of SC for assessing FEC, FB and complex SC in a sample of adolescent ASD and their NTD peers. We found that the use of tasks with fragmented emotional faces and the use of dynamic facial emotion stimuli lead to distinct differences in emotion categorization abilities between NTD and ASD.

In contrast, adolescents with ASD can solve classical FBT nearly as accurate as their NTD peers. We suggest that this is largely because of the simple structure of classical FBT, which enables adolescent individuals with ASD to find the correct answer by drawing inferences on explicitly given premises conveyed in verbal format instead of referring heavily on implicit ToM abilities that would be needed for example in non-verbal versions of FBTs.

In our study, especially SC tasks with a complex structure (like MASC and AToM) revealed impairments of SC in the ASD group. Such tasks are characterized by a plethora of relevant stimuli and all of them should be considered and handled to come to a correct evaluation of what has been perceived. In order to process this wealth of information quickly and implicitly, NTD draw on social schemata and scripts. We assume that impaired social cognition in ASD is attributable to deficiently developed social schemata. Thus, we proceed on the assumption that the deficient structure of social schemata is in part a result of the early emerging impairments in facial emotion categorization.

Limitations and Implications for Future Research

In this study we operationalized a theoretical approach via a wide range of various behavioral tests to investigate SC and the impact of iToM, eToM and socio-contextual abilities. As only large studies possess sufficient power to yield reliable results (e.g., Charman et al. 2011), our sample size is not big enough (i) to detect small and medium effects, (ii) to apply principal components analyses, or (iii) to perform more advanced statistical modelling approaches like SEM (Hair et al. 2014) that would be better suited for delineating between implicit, explicit, and complex social cognition. Thus, we suggest that future research should build upon larger sample sizes.

The age range (14–18 years) of our sample yielded significant group differences in implicit as well as contextual SC tasks. However, to make a clear statement about the trajectory of SC in terms of implicit, explicit and schema-based developments, it would be desirable to include younger cohorts as well as adults of all age groups.

Moreover, regarding the male-only sample, it is impossible to make any statement concerning gender differences in the development and abilities of social cognition.

Age and gender related aspects should also be reflected in the presentation of social stimuli. Stimuli of the DECT are limited to only one old male and one young female virtual character. Nevertheless, a wide range of cohorts and gendertypes could help to diversify the stimulus material and allow for valid conclusions about age and gender aspects of facial emotion categorization.

Our approach to investigate SC distinguishing between iToM and eToM with and without contextual factors is

operationalized by a behavioral test battery. To investigate underlying mechanisms and to understand the relevant processes applied in SC, further empirical and theoretical conceptualizations are needed.

In order to segregate the effect of contextual factors, future research should rely on consistent facial stimuli with and without contextual elements, to obtain distinct data allowing to deduct non-contextual from contextual factors.

The broad selection of our study tasks concerning different aspects of SC has revealed distinct impairments in ASD especially within the scope of FEC and complex social context. The findings point to a need to intensify some investigations that have not been elaborated sufficiently within this study.

According to the recommendations of Happé and Frith (2014) and the results of our study, we assume that implicit and explicit aspects of SC and the concept of social schemata should be investigated more thoroughly in terms of development. Opening the age range in both directions could be a remarkable chance to track whether and how attention to faces, emotion processing, facial emotion categorization, the ability to draw on contextual cues and the use of compensational heuristics, social schemata and scripts are related and in which chronological order they develop and appear.

As perceptual categorization of emotional stimuli seems to be considerably influenced by contextual factors and by contextual knowledge (Brosch et al. 2010), a profound analysis of the relationship between facial emotion categorization and additional sources of information could help to understand how the social brain network organizes the completion of insufficient information, especially in comparisons between ASD and NTD. In our study, we used diverse tests with different stimuli to investigate this phenomenon. Hence, the results and their implications should be replicated and tested by a task design dealing with consistent stimuli in a contextual and non-contextual condition.

The use of purely behavioral tasks for testing iToM-abilities in individuals with ASD compared to NTD participants often provides uncertain or ambiguous results in accuracy. Assisted by additional physiological measurements (eye-tracking, EEG, fMRI), most studies reveal deviant processing of the applied stimuli. For a better understanding of the perception, categorization and integration processes of SC especially in terms of contextual and non-contextual iToM and eToM tasks, it could be insightful to combine behavioral tests with eye-tracking, EEG or fMRI.

Further relevant methodological aspects to focus on are time constraints for the ecologically valid presentation of social stimuli. In our study only few tasks allowed for time limitations and if so, the time constraints were possibly not short enough to elicit significant differences between ASD and NTD. As we suggest that especially iToM is processed

mainly spontaneously, future studies should consider pertinent time-limited stimuli and reaction-time measurement in their design. Beside time limitations the use of dynamic stimuli provides higher ecological validity (Uljarevic and Hamilton 2012). Thus, future research should also focus on dynamic stimuli.

The data of the study presented here suggests that social cognition is a heterogeneous field comprising several different aspects and domains. Future research should operate, beside obligatory facial emotion categorization and false belief tasks, with ecologically valid material taking into account the complexity of the environment neurotypically developed and individuals with ASD are living in. Hence, a core theme of future research should be the investigation of differences in the acquisition and revision of social schemata between ASD and NTD.

Acknowledgments We would like to thank Volker Helzle, Simon Spielmann, Nils Zweiling and Diana Arellano of the Filmakademie Baden-Württemberg for the development of the virtual characters in the DECT and the deployment of Frapper Software. For providing stimulus material from the series *Shaun the Sheep*TM for the AToM test, we would like to thank Aardman Studios. For testing and data collection we would like to thank Annika Auch and Lara Seifert, Monica Biscaldi for expertise and helpful comments, Julian von Kügelgen, Gayatri Salunkhe, and Marie Vetter for proofreading and editing the article. Last but not least, we wish to thank three anonymous reviewers for their helpful comments and valuable suggestions.

Author contributions UMS conceptualized and designed the study, drafted the initial manuscript, helped to carry out the statistical analyses, reviewed and revised the manuscript, and approved the submission of the final manuscript. RR conceptualized and designed the study, supervised the initial manuscript, conducted the statistical analyses, helped to review and revise the manuscript, and approved the final manuscript for submission.

Compliance with Ethical Standards

Conflict of interest Ulrich Max Schaller and Reinhold Rauh declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants and their legal guardians.

Appendix

Example scene, questions and correct answers.

Example scene: The sheep deceives the farmer into abandon the swimming-pool, by letting him believe that clouds are gathering (see Fig. 1).

Question: iToM/Facial Emotion Expression (FEC): What facial expression does the farmer show when the sun disappears?

Correct Answers: angry, furious, irate.

Question: eToM/ False Belief 1st Order (FB1O): What does the farmer think, why the sun has disappeared?

Correct Answer: He thinks that clouds cover the sun.

References

- Aboulafia-Brakha, T., Christe, B., Martory, M.-D., & Annoni, J.-M. (2011). Theory of mind tasks and executive functions: A systematic review of group studies in neurology. *Journal of Neuropsychology*, 5(Pt 1), 39–55. doi:10.1348/174866410X533660.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological Review*, 116(4), 953–970. doi:10.1037/a0016923.
- Apperly, I. A., & Robinson, E. J. (2003). When can children handle referential opacity? Evidence for systematic variation in 5- and 6-year-old children's reasoning about beliefs and belief reports. *Journal of Experimental Child Psychology*, 85, 297–311. doi:10.1016/S0022-0965(03)00099-7.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men; research in human relations* (pp. 177–190). Oxford: Carnegie Press.
- Baron-Cohen, S. (2009). Autism: The empathizing-systemizing (E-S) theory. *Annals of the New York Academy of Sciences*, 1156, 68–80. doi:10.1111/j.1749-6632.2009.04467.x.
- Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21, 37–46.
- Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*, 29(5), 407–418. doi:10.1023/A:1023035012436.
- Baron-Cohen, S., Wheelwright, S., Hill, J. J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42(2), 241–251. doi:10.1111/1469-7610.00715.
- Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, 4(3), 311–331.
- Bekele, E., Crittendon, J., Zheng, Z., Swanson, A., Weitlauf, A., Warren, Z., & Sarkar, N. (2014). Assessing the utility of a virtual environment for enhancing facial affect recognition in adolescents with autism. *Journal of Autism and Developmental Disorders*, 44(7), 1641–1650. doi:10.1007/s10803-014-2035-8.
- Bodenhausen, G. V., Kang, S. K., & Peery, D. (2012). Social categorization and the perception of social groups. In S. T. Fiske & C. N. Macrae (Eds.), *SAGE handbook of social cognition* (pp. 311–329). Los Angeles, CA: Sage.
- Bölte, S., & Poustka, F. (2008). *SRS: Skala zur Erfassung sozialer Reaktivität. Dimensionale Autismus-Diagnostik. Deutsche Fassung der Social Responsiveness Scale (SRS) von John N. Constantino und Christian P. Gruber*. Bern: Huber.
- Bölte, S., Rühl, D., Schmötzer, G., & Poustka, F. (2006a). Diagnostisches Interview für Autistische Störungen revidiert (ADI-R). Deutsche Fassung des Autism Diagnostic Interview-revised. Bern: Huber.
- Bölte, S., Hubl, D., Feineis-Matthews, S., Prvulovic, D., Dierks, T., & Poustka, F. (2006b). Facial affect recognition training in autism: Can we animate the fusiform gyrus?. *Behavioral Neuroscience*, 120(1), 211–216. doi:10.1037/0735-7044.120.1.211.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.
- Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cognition and Emotion*, 24(3), 377–400. doi:10.1080/02699930902975754.
- Brosnan, M., Johnson, H., Grawmeyer, B., Chapman, E., & Benton, L. (2015). Emotion recognition in animated compared to human stimuli in adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 45(6), 1785–1796. doi:10.1007/s10803-014-2338-9.
- Carroll, J. M., & Russell, J. a (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, 70(2), 205–218.
- Cattell, R. R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1–22.
- Charman, T., Jones, C., Pickles, A., & Simonoff, E. (2011). Defining the cognitive phenotype of autism. *Brain Research*, 1380, 10–21.
- Christensen, W., & Michael, J. (2016). From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology*, 40, 48–64. doi:10.1016/j.newideapsych.2015.01.003.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. doi:10.1037/1040-3590.6.4.284.
- Clark, T. F., Winkielman, P., & McIntosh, D. N. (2008). Autism and the extraction of emotion from briefly presented facial expressions: Stumbling at the first step of empathy. *Emotion*, 8(6), 803–809. doi:10.1037/a0014124.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dawson, G., Webb, S. J., Carver, L., Panagiotides, H., & McPartland, J. (2004). Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Developmental Science*, 7(3), 340–359.
- Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*. doi:10.1111/cdev.12237.
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science*, 311(5763), 1005–1007. doi:10.1126/science.1121629.
- Döpfner, M., Plücker, J., Bölte, S., Lenz, K., Melchers, P., & Heim, K. (1998). *Fragebogen für Jugendliche; deutsche Bearbeitung der Youth Self-Report Form der Child Behavior Checklist (YSR). Einführung und Anleitung zur Handauswertung mit deutschen Normen*. Köln: Arbeitsgruppe Kinder- Jugendlichen- und Familiendiagnostik.
- Döpfner, M., Schmeck, K., & Berner, W. (1994). *Elternfragebogen über das Verhalten von Kindern und Jugendlichen. Forschungsergebnisse zur deutschen Fassung der Child Behavior Checklist (CBCL/4-18)*. Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD). Göttingen: Hogrefe.
- Dziobek, I., Fleck, S., Kalbe, E., Brand, M., Kessler, J., Woike, J. K., ... Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism*

- and *Developmental Disorders*, 36, 623–636. doi:10.1007/s10803-006-0107-0.
- Ekman, P., & Friesen, W. W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1), 56–75.
- Esteves, F., & Ohman, A. (1993). Masking the face: Recognition of emotional facial expressions as a function of the parameters of backward masking. *Scandinavian Journal of Psychology*, 34(1), 1–18.
- Evers, K., Steyaert, J., Noens, I., & Wagemans, J. (2015). Reduced recognition of dynamic facial emotional expressions and emotion-specific response bias in children with an autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 45(6), 1774–1784. doi:10.1007/s10803-014-2337-x.
- Farroni, T., Johnson, M. H., & Csibra, G. (2004). Mechanisms of eye gaze perception during infancy. *Journal of Cognitive Neuroscience*, 16(8), 1320–1326. doi:10.1162/0898929042304787.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal Psychology*, 58(2), 203–210.
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63, 287–313. doi:10.1146/annurev-psych-120710-100449.
- Frith, U., & Happé, F. (1994). Autism: Beyond “theory of mind”. *Cognition*, 50(1–3), 115–132.
- Gasper, K., & Clore, G. L. (2000). Do you have to pay attention to your feelings to be influenced by them? *Personality and Social Psychology Bulletin*, 26(6), 698–711. doi:10.1177/0146167200268005.
- Gehlen, A. (2014). *Der Mensch: Seine Natur und seine Stellung in der Welt (16.)*. Wiebelsheim: AULA-Verlag.
- Gepner, B., & Féron, F. (2009). Autism: A world changing too fast for a mis-wired brain? *Neuroscience and Biobehavioral Reviews*, 33(8), 1227–1242. doi:10.1016/j.neubiorev.2009.06.006.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–482. doi:10.1146/annurev-psych-120709-145346.
- Goldman, A. I. (2006). Simulating minds. In D. J. Chalmers (Ed.), *Philosophical books* (Philosophy, Vol. 49). New York: Oxford University Press. doi:10.1111/j.1468-0149.2008.459_15.x.
- Grossmann, T., Johnson, M. H., Farroni, T., & Csibra, G. (2007). Social perception in the infant brain: Gamma oscillatory activity in response to eye gaze. *Social Cognitive and Affective Neuroscience*, 2(4), 284–291. doi:10.1093/scan/nsm025.
- Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., & Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1653), 2803–2811. doi:10.1098/rspb.2008.0986.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th edn.). Harlow: Pearson Education Limited.
- Happé, F. G. E. (1994). 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*, 24(2), 129–154. doi:10.1007/BF02172093.
- Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66(3), 843–855.
- Happé, F., & Frith, U. (2014). Annual research review: Towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology and Psychiatry*, 55(6), 553–577. doi:10.1111/jcpp.12162.
- Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, 20(3), 290–322. doi:10.1007/s11065-010-9138-6.
- Heatherton, T. F., Macrae, C. N., & Kelley, W. M. (2004). What the social brain science can tell us about the self? *Current Directions in Psychological Science*, 13(5), 190–193.
- Helzlsouer, V., Biehn, C., Schlömer, T., & Linner, F. (2004). Adaptable setup for performance driven facial animation. *ACM SIGGRAPH 2004*.
- Hertwig, R., & Hoffrage, U. (2013). *Simple heuristics in a social world*. New York: Oxford University Press.
- Higgins, E. (1998). The aboutness principle: A pervasive influence on human inference. *Social Cognition*, 16(1), 173.
- Hill, E. L., Sally, D., & Frith, U. (2004). Does mentalising ability influence cooperative decision-making in a social dilemma? Introspective evidence from a study of adults with autism spectrum disorder. *Journal of Consciousness Studies*, 11(7), 144–161.
- Hoernig, R., Rauh, R., & Strube, G. (1993). EVENTS-II: Modeling event recognition. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 113–138). Amsterdam: North-Holland.
- Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns’ preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1–2), 1–19.
- Johnson-Laird, P. (1981). *Mental models in cognitive science*. *Cognitive Science* (Vol. 4). doi:10.1016/S0364-0213(81)80005-5.
- Jolliffe, T., & Baron-Cohen, S. (1999). The strange stories test: A replication with high-functioning adults with autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*, 29(5), 395–406. doi:10.1023/A:1023082928366.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., & Frederick, S. (2002). *Representativeness revisited: Attribute substitution in intuitive judgment*. *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.
- Kessels, R. P. C., Spee, P., & Hendriks, A. W. (2010). Perception of dynamic facial emotional expressions in adolescents with autism spectrum disorders (ASD). *Translational Neuroscience*, 1(3), 228–232. doi:10.2478/v10134-010-0033-8.
- Kessler, H., Bayerl, P., Deighton, R. M., & Traue, H. C. (2002). Facially expressed emotion labeling (FEEL): A computer test for emotion recognition. *Verhaltenstherapie und Verhaltensmedizin*, 23(3), 297–306.
- Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport*, 14(18), 2317–2322. doi:10.1097/01.wnr.0000101520.44335.20.
- Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., & Whalen, P. J. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of Cognitive Neuroscience*, 16(10), 1730–1745. doi:10.1162/0898929042947865.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394. doi:10.1037/0033-295X.85.5.363.
- Kliemann, D., Rosenblau, G., Bölte, S., Heekeren, H. R., & Dziobek, I. (2013). Face puzzle-two new video-based tasks for measuring explicit and implicit aspects of facial emotion recognition. *Frontiers in Psychology*, 4, 376. doi:10.3389/fpsyg.2013.00376.
- Kupfer, J., Brosig, B., & Brähler, E. (2001). TAS-26: Toronto Alexithymie-Skala-26. Göttingen: Hogrefe.
- Leppanen, J. M., & Hietanen, J. K. (2003). Affect and face perception: Odors modulate the recognition advantage of happy faces. *Emotion*, 3(4), 315–326. doi:10.1037/1528-3542.3.4.315.

- Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion, 14*(2), 375–387. doi:10.1037/a0035293.
- Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *The British Journal of Developmental Psychology, 30*(Pt 1), 1–13. doi:10.1111/j.2044-835X.2011.02074.x.
- Lozier, L. M., Vanmeter, J. W., & Marsh, A. a. (2014). Impairments in facial affect recognition associated with autism spectrum disorders: A meta-analysis. *Development and Psychopathology, 26*, 1–13. doi:10.1017/S0954579414000479.
- Macrae, C. N., & Miles, L. K. (2012). Revisiting the sovereignty of social cognition: Finally some action. In S. T. Fiske & C. N. Macrae (Eds.), *The SAGE handbook of social cognition* (pp. 1–11). London: SAGE Publications.
- Marchi, F., & Newen, A. (2015). Cognitive penetrability and emotion recognition in human facial expressions. *Frontiers in Psychology, 6*, 1–12. doi:10.3389/fpsyg.2015.00828.
- Martinez, A., & Du, S. (2012). A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *The Journal of Machine Learning Research, 13*(13), 1589–1608.
- Matsumoto, D., & Ekman, P. (1988). Japanese and Caucasian facial expressions of emotion (JACFEE) [Slides]. San Francisco, CA: Intercultural and Emotion Research Laboratory, Department of Psychology, San Francisco State University.
- McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlations coefficients. *Psychological Methods, 1*(1), 30–46.
- Mehu, M., Grammer, K., & Dunbar, R. I. M. (2007). Smiles when sharing. *Evolution and Human Behavior, 28*(6), 415–422. doi:10.1016/j.evolhumbehav.2007.05.010.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal Psychology, 67*(4), 371–378.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howarter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. doi:10.1006/cogp.1999.0734.
- Murphy, N. A. (2012). Nonverbal perception. In S. T. Fiske & C. N. Macrae (Eds.), *The SAGE handbook of social cognition* (1., pp. 191–210). London: SAGE Publications.
- Oikawa, M. (2010). Unconscious thinking, doing, and wanting: A review of mechanisms and moderators of assimilation and contrast effects. *Japanese Psychological Review, 53*(4).
- Parish-Morris, J., Chevallier, C., Tonge, N., Letzen, J., Pandey, J., & Schultz, R. T. (2013). Visual attention to dynamic faces and objects is linked to face processing skills: A combined study of children with autism and controls. *Frontiers in Psychology, 4*, 185. doi:10.3389/fpsyg.2013.00185.
- Payne, B. K. (2012). Control, awareness, and other things we might learn to live without. In S. T. Fiske & C. N. Macrae (Eds.), *The SAGE handbook of social cognition*. (pp. 12–31). London: SAGE Publications. doi:10.4135/9781446247631.
- Pelphrey, K. A., Morris, J. P., McCarthy, G., & Labar, K. S. (2007). Perception of dynamic changes in facial affect and identity in autism. *Social Cognitive and Affective Neuroscience, 2*(2), 140–149. doi:10.1093/scan/nsm010.
- Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*(9), 337–344.
- Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that ...” attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437–471.
- Portmann, A. (1941). Die biologische Bedeutung des ersten Lebensjahres beim Menschen. *Schweizerische Medizinische Wochenschrift, 71*(1941), 921–924.
- Rahko, J. S., Paakki, J.-J., Starck, T. H., Nikkinen, J., Pauls, D. L., Kätsyri, J. V., ... Kiviniemi, V. (2012). Valence scaling of dynamic facial expressions is altered in high-functioning subjects with autism spectrum disorders: An fMRI study. *Journal of Autism and Developmental Disorders, 42*(6), 1011–1024. doi:10.1007/s10803-011-1332-8.
- Rauh, R., & Schaller, U. M. (2009). *Categorical perception of emotional facial expressions in video clips with natural and artificial actors: A pilot study* (Technical Report ALU-KJPP-2009-001). Freiburg: Research Group for Developmental Cognitive Psychiatry, Department of Child and Adolescent Psychiatry and Psychotherapy, University of Freiburg.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- Rigato, S., Menon, E., & Johnson, M. H. (2011). Direct gaze may modulate face recognition in newborns. *Infant and Child Development, 34*, 20–34. doi:10.1002/icd.
- Righart, R., & de Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. *Social Cognitive and Affective Neuroscience, 3*(3), 270–278. doi:10.1093/scan/nsn021.
- Rochat, P., Querido, J. G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology, 35*(4), 950–957.
- Rühl, D., Bölte, S., Feineis-Matthews, S., & Poustka, F. (2004). Diagnostische Beobachtungsskala für Autistische Störungen (ADOS). Deutsche Fassung der Autism Diagnostic Observation Schedule. Bern: Huber.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*(1), 60–94. doi:10.1037/0033-295X.89.1.60.
- Rump, K. M., Giovannelli, J. L., Minschew, N. J., & Strauss, M. S. (2009). The development of emotion recognition in individuals with autism. *Child Development, 80*(5), 1434–1447. doi:10.1111/j.1467-8624.2009.01343.x.
- Sasson, N. J., Pinkham, A. E., Weittenhiller, L. P., Faso, D. J., & Simpson, C. (2015). Context effects on facial affect recognition in schizophrenia and autism: Behavioral and eye-tracking evidence. *Schizophrenia Bulletin, 42*(3), 1–9. doi:10.1093/schbul/sbv176.
- Sato, W., Uono, S., & Toichi, M. (2013). Atypical recognition of dynamic changes in facial expressions in autism spectrum disorders. *Research in Autism Spectrum Disorders, 7*(7), 906–912. doi:10.1016/j.rasd.2013.04.008.
- Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences, 19*(2), 65–72. doi:10.1016/j.tics.2014.11.007.
- Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69*(5), 379–399. doi:10.1037/h0046234.
- Schaller, U. M., & Rauh, R. (2011). *Animated-Theory-of-Mind (AToM): A Test for Measuring Implicit and Explicit Theory of Mind Taking Account of Dynamic and Contextual Aspects* (Technical Report ALU-KJPP-2011-001). Freiburg: Research Group for Developmental Cognitive Psychiatry, Department of Child and Adolescent Psychiatry and Psychotherapy, University of Freiburg.

- Schank, R. C., & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding - An inquiry into human knowledge structures*. Hillsdale: Erlbaum.
- Scharlemann, J., Eckel, C., Kacelnik, A., & Wilson, R. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, 22(5), 617–640. doi:10.1016/S0167-4870(01)00059-9.
- Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General*, 141(3), 433–438. doi:10.1037/a0025458.
- Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition*, 133(1), 43–47. doi:10.1016/j.cognition.2014.05.016.
- Schuerk, T., Vuori, M., & Sodian, B. (2014). Implicit and explicit Theory of Mind reasoning in autism spectrum disorders: The impact of experience. *Autism: The International Journal of Research and Practice*, 19, 459–468. doi:10.1177/1362361314526004.
- Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *Neuroscientist*, 18(2), 108–113. doi:10.1177/1073858410397208.
- Senju, A. (2013). Atypical development of spontaneous social cognition in autism spectrum disorders. *Brain and Development*, 35(2), 96–101. doi:10.1016/j.braindev.2012.08.002.
- Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., ... Csibra, G. (2010). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology*, 22(2), 353–360. doi:10.1017/S0954579410000106.
- Sherman, J. W., Gawronski, B., & Trope, Y. (2014). *Dual-process theories of the social mind*. New York: Guilford Publications.
- Shi, H., Wang, X., & Yao, S. (2013). Comparison of activation patterns between masking and inattention tasks: A coordinate-based meta-analysis of implicit emotional face processing. *Frontiers in Human Neuroscience*, 7, 459. doi:10.3389/fnhum.2013.00459.
- Soto-Icaza, P., Aboitiz, F., & Billeke, P. (2015). Development of social skills in children: Neural and behavioral evidence for the elaboration of cognitive models. *Frontiers in Neuroscience*, 9, 1–16. doi:10.3389/fnins.2015.00333.
- Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General*, 141(4), 715–727. doi:10.1037/a0027950.
- Sugranyes, G., Kyriakopoulos, M., Corrigall, R., Taylor, E., & Frangou, S. (2011). Autism spectrum disorders and schizophrenia: Meta-analysis of the neural correlates of social cognition. *PLoS One*, 6(10), e25322. doi:10.1371/journal.pone.0025322.
- Todorov, A. (2012). The social perception of faces. In S. T. Fiske & C. N. Macrae (Eds.), *The SAGE handbook of social cognition* (pp. 96–114). London: SAGE Publications.
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, 27(6), 813–833. doi:10.1521/soco.2009.27.6.813.
- Tracy, J. L., Robins, R. W., Schriber, R. A., & Solomon, M. (2010). Is emotion recognition impaired in individuals with autism spectrum disorders? *Journal of Autism and Developmental Disorders*, 41(1), 102–109. doi:10.1007/s10803-010-1030-y.
- Trillingsgaard, A. (1999). The script model in relation to autism. *European Child and Adolescent Psychiatry*, 8(1), 45–49. doi:10.1007/s007870050082.
- Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, 59, 329–360. doi:10.1146/annurev.psych.59.103006.093707.
- Uljarevic, M., & Hamilton, A. (2012). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, 43(7), 1517–1526. doi:10.1007/s10803-012-1695-5.
- Uono, S., Sato, W., & Toichi, M. (2009). Dynamic fearful gaze does not enhance attention orienting in individuals with Asperger's disorder. *Brain and Cognition*, 71(3), 229–233. doi:10.1016/j.bandc.2009.08.015.
- Van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, 108(3), 796–803. doi:10.1016/j.cognition.2008.07.002.
- Vogele, K. (2009). Soziale Kognition und soziales Bewusstsein. In M. Jung (Ed.), *Funktionen des Erlebens: Neue Perspektiven des qualitativen Bewusstseins* (pp. 285–304). Berlin: de Gruyter.
- Weiß, R. H. (2008). *CFT20-R. Grundintelligenzskala 2 - revision*. Göttingen: Hogrefe.
- 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.
- Whalen, P. J., Rauch, S. L., Etkoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, 18, 411–418.
- Wichmann, S. (1996). *Wie Kinder über das Denken denken*. Frankfurt am Main: Peter Lang.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7), 592–598.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.
- Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, 5(3), 277–283. doi:10.1038/nm816.
- Yamasaki, H., LaBar, K. S., & McCarthy, G. (2002). Dissociable prefrontal brain systems for attention and emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17), 11447–11451. doi:10.1073/pnas.182176499.