

Action Understanding in High-Functioning Autism: The Faux Pas Task Revisited

Joanna Korman (joanna.korman.ctr@nrl.navy.mil)

Navy Center for Applied Research in Artificial Intelligence
Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC 20375

Tiziana Zalla (tiziana.zalla@ens.fr)

Institut Jean Nicod-CNRS, Pavillon Jardin
Ecole Normale Supérieure, 29, rue d'Ulm, 75005, Paris, France

Bertram F. Malle (bertram_malle@brown.edu)

Department of Cognitive, Linguistic, & Psychological Sciences
Brown University, 190 Thayer St. Providence, RI 02912

Abstract

Individuals with autism spectrum disorders (ASD) are said to have deficits in “theory of mind.” The present paper explores two main accounts of the mechanisms underlying these deficits. On one account, high-functioning adults with ASD struggle to infer others’ mental states. On another account, they lack an ability to *integrate* those mental states into a coherent understanding of action. We tested these two accounts by making several modifications to the Faux Pas task—a commonly used advanced theory of mind task—including the presentation of explicit mental state information. Surprisingly, in contrast to previous work, individuals on the autism spectrum exhibited both intact integration and intact inference.

Keywords: Theory of mind; intentional action; autism spectrum disorder; mental state inference

Autism and Theory of Mind: Belief and Action Understanding

Autism spectrum disorder (ASD) is a developmental disorder characterized by deficits in reciprocal social interaction and communication (DSM-V, American Psychiatric Association, 2013). In contrast to typically developing children, autistic children are widely described as having deficits in “Theory of Mind” (ToM), or the ability to represent the mental states of other people. Characteristic of these deficits have been ASD children’s early failures to pass false belief tasks at the normative age of four years (Baron-Cohen, 1985). But by the time they reach adulthood, many high-functioning adults on the autism spectrum succeed at traditional false belief tasks (Bowler, 1992). Because ASD individuals require higher verbal ability than typically developing adults to pass such tasks, researchers have proposed that ASD individuals do so by using deliberate, conscious calculation (Happé, 1995) and by relying on specific features of language, such as complement syntax (Lind & Bowler, 2009).

However, high-functioning autistic adults who pass classic ToM tasks persist in their social deficits (e.g., Klin, Jones, Schultz, & Volkmar, 2003). Therefore, such tasks cannot be capturing these social deficits’ core features.

Hence, autism researchers have developed more “advanced” ToM tasks to highlight the persistence of broader theory of mind deficits in more naturalistic settings. Although researchers have succeeded at demonstrating that individuals on the autism spectrum struggle with these novel tasks (Baron-Cohen, 1999; Zalla, Sav, Stopin, Ahade & Leboyer, 2009), the precise mechanisms underlying these struggles have gone largely unexplored. The present paper explores these mechanisms by examining a modified version of one such task.

The Faux Pas Task: Revealing Deficits in Adults with ASD

The “faux pas” task (Baron-Cohen, O’Riordan, & Jones, 1999; Zalla et al., 2009) presents a context in which one character (the speaker) makes a statement that is unintentionally offensive to the listener because the speaker has a false belief. For example, in one story, Jane moves into a new apartment and purchases new curtains for the windows. When her best friend Lisa comes over, she says to Jane, “Oh, I hope you’re going to get new curtains! These ones are awful!” Typically developing individuals recognize that Lisa’s comment is offensive to Jane, but when asked why Lisa said that, infer that Lisa did not know the curtains were picked out by Jane herself. In contrast, while individuals with ASD can detect that something was “wrong” or “awkward” with Lisa’s comment, they struggle to detect that Lisa made the comment unintentionally – that she had a false belief and made the statement out of a positive or neutral desire (e.g., to be helpful with decorating). The outcome is not just an unfortunate side effect of an otherwise fulfilled intention; the complete falsity of the agent’s belief actually precludes the fulfillment of the speaker’s desire. Unlike controls, adults with ASD demonstrated mixed success in detecting the character’s false belief and positive desire. They sometimes acknowledged that the speaker had a positive desire but often failed to correctly infer the speaker’s belief state. And in a small number of cases (10% of all responses), they even incorrectly attributed a negative intention to the speaker

(e.g., Lisa wants to insult Jill’s taste in décor) (Zalla et al., 2009).

The Faux Pas Task: What Does it Measure?

What explains the struggles that high-functioning individuals with ASD encounter on these tasks? The faux pas task has previously been described as a more naturalistic and more robust way than ‘false belief’ tasks of measuring mental state understanding. But understanding the mental states behind the kinds of actions presented in the Faux Pas task can be broken down into several distinct subtasks, outlined below and in Figure 1.

Conceptual integration First, achieving a holistic understanding of the depicted complex behavior requires an understanding of the *concept of intentional action*. This concept requires the social perceiver to grasp more than just individual mental state concepts of belief and desire: rather, it requires understanding how these individual concepts are *integrated* to support the understanding of an action as intentional. That is, an action is performed *intentionally* only if the agent had a desire for the action’s outcome and a belief that her action would lead to (serve as a means to achieving) that particular outcome (Malle & Knobe, 1997). In the present example, for instance, understanding what action was intentional (making a certain remark) and what outcomes were unintentional (the remark offended Jane) requires understanding both that Lisa had a false belief (that the curtains were bought by the previous owner) and that she wanted to be critical only of the previous owner, and that she therefore could not have offended Jane intentionally.

Mental state inference In addition to requiring participants to have a concept of intentional action, the faux pas task requires another capacity: the ability to *infer* the specific contents of the character’s mental states. For example, it is not explicitly stated in the story that Lisa has a false belief about who bought the curtains or that she wanted to be critical of the previous owner; the social perceiver must infer this belief and this desire from the story. Further, there are at least two different mechanisms by which individuals with ASD could be failing to correctly infer the characters’ mental states (see Figure 1). On the consensus account, the faux pas task is simply an “advanced” way of revealing enduring deficits in theory of mind – the ability to generate the contents of specific mental states, most particularly, belief – that are already revealed in simpler false belief tasks earlier in life. We refer to this account as the enduring “**theory of mind**” deficit account.

On an alternate account, however, individuals may struggle on the faux pas task not because they have an enduring deficit in generating the contents of mental states, but because the faux pas task contains an increase in complexity over traditional false belief tasks, thereby disallowing the use of familiar (linguistic) compensation strategies. For example, individuals with ASD may struggle on the faux pas task because this task presents stories in which it is necessary to make a rich suite of background

assumptions about the agent’s social roles and context (e.g., the idea that friends normally do not insult their friends’ curtains). Such a task does not allow for the use of simple rules (such as “perception leads to knowledge”) to generate mental state contents; it instead requires social perceivers to draw productively on their general knowledge to produce accurate inferences. Since individuals with ASD also have documented deficits in this type of knowledge-based inferential generativity (Loth, Gómez, & Happé, 2008), it is possible that struggles on the faux pas task are due to a general **knowledge-based inference deficit**, and not any deficit in theory of mind per se.

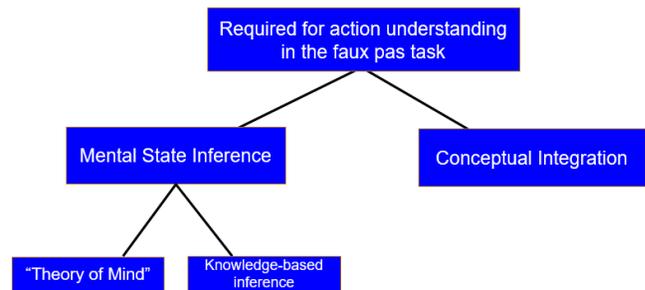


Figure 1. Capacities required for action understanding in the Faux Pas task.

The Present Study: Which Hypothesis Explains Low Performance?

To distinguish between the *mental state inference hypothesis* and the *conceptual integration* (or intentional action understanding) *hypothesis*, we developed 8 novel vignettes based on the faux pas task (Zalla et al., 2009; Baron-Cohen et al., 1999) and created four conditions of varying information availability. In the “No Information” condition, we presented participants with no explicit mental state information, as in the original Faux Pas task. In three additional conditions, we presented them with either the character’s belief only (Belief condition), the character’s desire only (Desire condition), or both the character’s belief and desire (Full Information condition). A sample story in the Full Information formulation follows. In the “No Information” condition, the underlined text would be omitted:

Clara is very short and dresses plainly. One day she goes to pick up her son James from school early for a medical appointment. Clara enters the school and spots James’s teacher, Mrs. Hayes. Mrs. Hayes thinks that Clara is a student lost in the hallway. [Belief] Mrs. Hayes wants to help [Desire]. Before Clara can ask after James’ whereabouts, Mrs. Hayes looks at Clara and says, “Have you lost your class, honey?”

Mental State Inference vs. Conceptual Integration By comparing stories in the three explicit mental state information conditions with the No Information condition, we can broadly distinguish between the inference and

integration hypotheses. If adults with ASD are capable of conceptually integrating mental state information for action understanding but are not capable of inferring this information, they should show improved performance in the presence of explicit mental state information over the “No Information” case. In contrast, if they are not capable of integrating mental state information for intentional action understanding, they should struggle to accurately interpret the meaning of the agents’ actions as depicted in the stories even when mental state information is explicitly presented to them.

Two Mechanisms of Mental State Inference In addition to distinguishing broadly between mental state inference and conceptual integration, we also sought to distinguish two possible mechanisms of mental state inference deficit: the commonly cited theory of mind deficit and a nonspecific general knowledge deficit. In previous versions of the Faux Pas task, pieces of general knowledge (such as the fact that, in the above sample story, Clara probably appears to be younger than she really is) were not included in the story and instead had to be inferred – *in addition* to mental state information. Low performance on the task may thus have been a result of failed general knowledge inferences, not a result of failed ToM inferences. In the present task, we sought to mitigate this ambiguity by explicitly providing such background information in every condition (e.g., “Clara is very short and dresses plainly”) and requiring participants to infer only the missing mental state information (e.g., that, in light of Clara’s appearance, Mrs. Hayes falsely believed Clara to be a student). If participants continue to struggle to produce belief inferences in the No Information condition in spite of these additions of inference-ready background information, then we can be confident that the present faux pas stories indeed measure only mental state inference (or “theory of mind”) capabilities, and not any additional abilities.

Conceptual Integration Apart from inference, by explicitly presenting mental state information, we can test ASD individuals’ abilities to integrate this information in the service of intentional action understanding. To demonstrate integration capacity, individuals on the autism spectrum must go beyond understanding *the fact that* Lisa had a false belief. They must also show understanding of *how* her false belief is *relevant* to the action and its outcome: that it was *because* she had a false belief that she made the remark, and that the remark resulted in offense *because* of that false belief (Lisa didn’t realize that it would lead to a negative outcome when she said it) (Figure 2). Similarly, understanding *that* Lisa had a positive desire is not sufficient for a full conceptual understanding of the action. In the presence of a positive desire, individuals with ASD must recognize that the action (utterance) still may have *caused* a negative outcome, even though the desire motivating it was a positive one. In summary, to show

integration capacity, individuals with ASD must be able to see how a story character’s mental states relate to her action and its outcomes.

Most centrally, in the presence of both a belief and a desire (Full Information condition), the ability to correctly understand that the action is intended to be a positive one, was caused by the story character’s false belief, and has a negative outcome, demonstrates intact integration capacities. In addition, the inclusion of conditions in which either a belief or a desire alone was presented allowed us to test integration abilities under somewhat more difficult conditions.¹

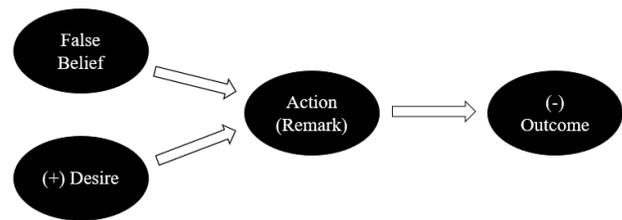


Figure 2. Conceptual integration requires the social perceiver to recognize the relevance of the story character’s mental states to the action’s meaning and outcome.

Study

Procedure and Measures

We presented control and ASD participants with eight faux pas stories each: two of each in each of the four mental state information conditions (No Information, Belief only, Desire only, Full Information). Participants also received six control stories. Participants read each story and then answered several forced-choice and open-ended questions about the story (detailed below), which served both as measures of inference and integration depending on information condition: in the No Information condition, measures of belief understanding, explanations for the action, and description of the action were a measure of inferential abilities, and in the explicit mental state information condition, these same measures demonstrated participants’ abilities to integrate provided mental states into a coherent understanding of action.

Participants 20 participants with Autism Spectrum Disorder (ASD), as confirmed on the ADOS-2 (Lord, Rutter, DiLavore, Risi, Gotham, & Bishop, 2012), were recruited in partnership with the Rhode Island Consortium for Autism Research and Treatment (RI-CART), $M_{Age} = 31.90$ years; 5 Female; mean score on Ravens Progressive Matrices, 9-item short form intelligence test (Bilker et al., 2012), $M = 45.94/60$, $SD = 8.0$. 20 typically developing controls were

¹ Each of these two single-mental state conditions involves more specific tests of integration capacity. For example, in explicit desire (alone) condition, the participant’s ability to produce a false belief indicates that she can recognize the relevance of the provided desire to an action that produces an unforeseen outcome.

recruited to match the ASD group on age, gender, and intelligence ($M_{Age} = 30.35$; 6 Female; $M_{Ravens} = 49.75$, $SD = 14.29$).

Interpretation of the Utterance After reading each story, participants were first prompted to describe the main character’s utterance. They were instructed to “check all that apply” among four options: “It was awkward,” “It was nice,” “It was mean,” and “It was neutral.”

Belief Question Participants answered whether the character who made the utterance possessed a false belief (e.g., “Did Mrs. Hayes believe that Clara was James’s mother?”)

Explanation Question Participants then answered in a text box the “explanation” question, which simply asked, “Look back at what [character] said. Why did [s/he] say that?” These open-ended responses were content coded for explanatory quality. Two coders classified each response into a single numbered category, 0-3. To receive a perfect score of 3, the participant had to give an explanation that directly stated or otherwise implied that the speaker’s action was caused by that character’s false belief. Incorrect responses, such as those that cited a negative intention on the part of the speaker, received a score of “0”.

Results and Discussion

Interpretation of the Utterance To achieve a data reduction of the 32 cells represented (4 descriptors for each of 4 information conditions, rated by two different participant groups), responses for each of the four variables were first aggregated across each cell of the 2 (autism vs. control) X 4 (mental state information) design. Values on each variable were aggregated across participant group and information condition, yielding one variable for each descriptor, and entered into a principal components analysis. Two orthogonal components were extracted. Component loadings for each of the four variables are provided in Table 1.

Table 1. Principal components analysis of four utterance interpretation options results in two components

Descriptor	Unintentional	Motive
Awkward	0.65	0.33
Nice	-0.48	0.35
Mean	-0.32	-0.61
Neutral	-0.08	0.45

The two components can be best interpreted as capturing (1) whether the utterance was classified as awkwardly *unintentional* (awkward) vs. intentional (nice or mean) and (2) whether the *motive* behind the action was classified as nice (or neutral) vs. mean. Component scores for each participant were computed from linear combinations of the

four constituent variables, and ranged in value from -2 to 2. A score close to +2 on the Unintentional component indicated that “awkward” was checked for both stories in that condition (and that the slight was therefore unintentional), while a score close to -2 indicated that the utterance in that story had instead been classified as intentional. On the Motive component, scores closer to +2 indicated less endorsement of the “mean” descriptor and more endorsement of the “nice” or “neutral” descriptors.

For this and subsequent measures, we performed two main analyses. To test the inference hypothesis, we compared the two participant groups with t tests in the No Information condition only. To test the integration hypothesis, we performed ANOVAs on the four conditions for both participant groups. Of primary interest to the integration hypothesis were interactions between participant group (Autism vs. Control) and information condition (primarily, the three explicit information conditions vs. the No Information condition).

On the Unintentional component (Figure 3), controls recognized the utterance as “awkward” with greater frequency in the No Information condition than did ASD participants, $t(34.72) = 2.04$, $p < .05$. In addition, there was an interaction between Autism and Information condition, but it did not reach significance, $F(1, 38) = 3.14$, $p = .08$. On the Motive component (Figure 4), there were no group differences in the No Information condition, nor were there significant main effects or interactions of information condition with participant group.

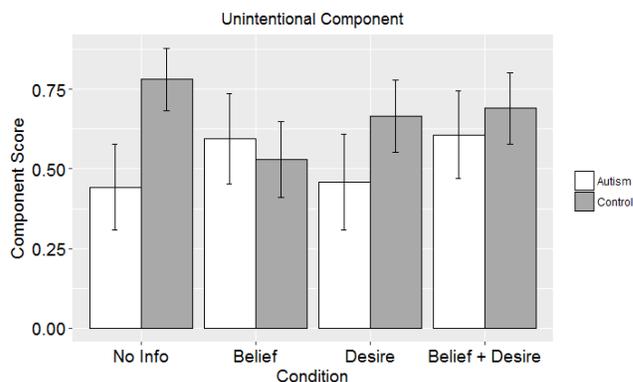


Figure 3: Unintentional Component

Implications for Inference Hypothesis Compared to control participants, ASD participants were less capable of recognizing that the speaker’s utterance could be described as awkward. However, the two groups did not differ in any of the explicit mental state information conditions with respect to this recognition, suggesting that the no-information difference may be spurious or not due to a deficit on the part of ASD participants. When assessing the possible motives of the speaker, ASD participants exhibited no difficulties identifying the positive desire (lack of ‘mean’ intent) underlying the speaker’s utterance.

Implications for Integration Hypothesis In spite of some differences with controls in the No Information condition, ASD individuals performed similarly to controls in the presence of explicit mental state information, identifying the utterance’s awkwardness with similar frequency and the story character’s positive desire with similar frequency.

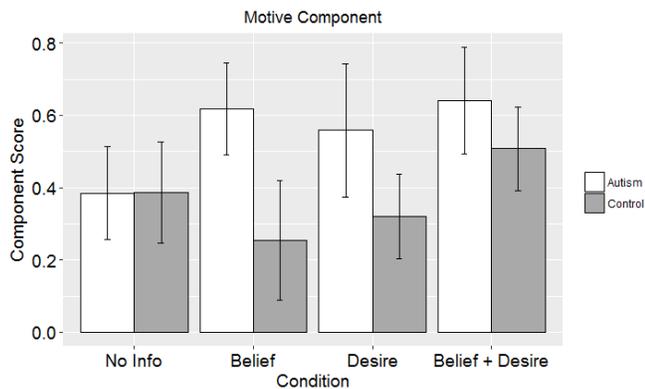


Figure 4: Motive Component

Belief Question Correct responses to the belief question were aggregated across the two stories comprising each of the four information conditions, yielding a score of 0 to 2 per condition (see Figure 5).

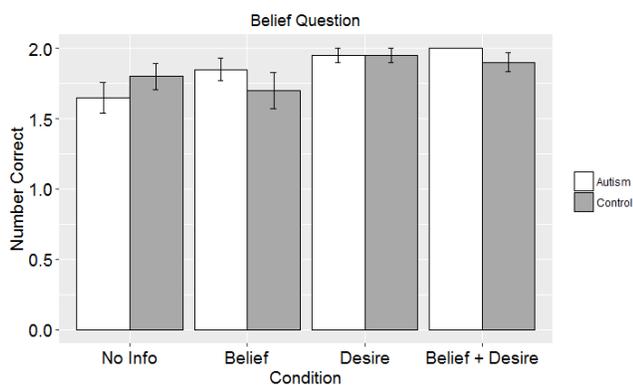


Figure 5: Belief Question

To examine the inference hypothesis, we compared performance of ASD participants ($M = 1.65$ stories correct) and Control participants ($M = 1.80$) in the No Information condition. There was no difference in performance, Welch’s $t(36.88) = 1.05, p = 0.30$.

To test the integration hypothesis, we examined whether adding mental state information affected the accuracy of correctly recognizing the belief, again performing a 4 (Information condition) X 2 (Autism vs. Control) mixed ANOVA. There was a main effect of information condition: Participants as a whole exhibited improved performance in the presence of explicit mental state information (in the three explicit mental state information conditions) as compared to the No Information condition, $F(1, 38) = 4.36, p = .04$. No interactions between Autism and Information conditions reached significance.

Implications for Inference Hypothesis In our No Information condition, all general knowledge was made explicit, and ASD participants performed as well as controls. By contrast, in comparable no-information conditions, previous Faux pas studies did not make such background knowledge available and documented performance decrements for ASD participants. This pattern of results supports the hypothesis that previously documented inferential deficits may not be due to deficits in “theory of mind” per se, but may instead have been caused by the inability to draw inferences from general knowledge.

Implications for Integration Hypothesis Neither participants with ASD nor control participants performed at ceiling in the No Information condition, leaving room for improvement (and a demonstration of integration capacities) in the presence of explicit mental state information. Although there was a main effect of information condition, there was no difference between ASD and control participants. Thus, ASD participants appear to be as capable as control participants at integrating explicitly presented belief and desire information.

Explanation Question Once more, there was no difference between ASD and control participants in the No Information condition, Welch’s $t(36.52) = .40, p = .70$. We also found no main effects or interactions involving the comparison between ASD and control participants across all four conditions (all $ps > 0.36$). There were, however, significant main effects of information condition, with all participants providing higher-quality explanations when receiving explicit information about both mental states than in the No Information condition, $F(38) = 3.71, p = .06$, and higher-quality explanations when receiving explicit information about both a belief and desire than when receiving information about either one of these mental states alone, $F(38) = 4.44, p < .05$.

Implications ASD and control participants provided equally accurate explanations in all information conditions, including the most challenging one (where no explicit belief or desire information was provided). Moreover, like control participants, ASD participants improved their explanation quality in response to explicit mental state information, suggesting in particular that integration capacities held by control participants are also held by those with ASD. Thus, we may conclude once more that previously documented deficits for ASD individuals – in both inference and integration – may have been caused by other aspects of the task, such as the requirement of general knowledge recruitment.

General Discussion

We considered two main hypotheses that could explain deficits for individuals with ASD in demanding theory of mind tasks such as the Faux Pas task. One suggests that individuals with autism struggle to generate the contents of mental states (inference), while the other suggests that they

struggle to integrate mental state information to reach a full understanding of action (integration). Both hypotheses fall short of explaining our data.

Mental State Inference Deficits?

Contrary to previous studies of the Faux Pas task (Zalla et al. 2009, Baron Cohen et al., 1999) as well as other similar advanced theory of mind tasks (e.g., Happé, 1994), ASD participants in our study performed comparably to control participants, even when receiving no explicit mental state information. This performance spanned a number of measures, including correct inferences of the story character's false belief and positive desire. It appears that, in the presence of enriched background information to afford inferences from general knowledge, ASD participants more capably inferred mental states than they did in previous studies in which stimuli lacked such enriched background information. Although the present study did not directly compare background-enriched stories with unenriched stories, this finding is suggestive: previously documented deficits on advanced theory of mind tasks may depend on a suite of inferential capacities, of which mental state inference, per se, or "theory of mind," is only one, and perhaps a less influential one.

Integration Deficits?

In addition to demonstrating intact inferential abilities in the presence of enriched background information, ASD participants in our study also demonstrated intact abilities to integrate provided mental state information into a coherent understanding of intentional action. Even in response to a challenging, open-ended question about the character's utterance—"Why did he say that?"—participants with ASD accurately linked mental states with action as well as controls did.

With the addition of (1) enriched background information and (2) explicit mental state information, high-functioning adults with ASD exhibited a remarkable ability to comprehend the meaning behind a story character's complex intentional action. This finding is notable in light of previous work suggesting that individuals with ASD struggle to reach the requisite mental state inferences (Baron Cohen et al. 1985, Happé 1994), and to integrate a character's mental states with her action's outcome to reach a full comprehension of that action (Moran, Young, Saxe, Lee, O'Young, Mavros, & Gabrieli, 2011).

Acknowledgments

Thanks to Corey Cusimano, Baxter DiFabrizio, Stuti Thapa Magar, Lindsay Oberman, Stephen Sheinkopf, and all the staff of the Rhode Island Consortium for Autism Research and Treatment (RICART).

References

American Psychiatric Association. (2013). Diagnostic and statistical manual (5th ed). Washington, D.C.

- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1),
- Baron-Cohen, S., O'Riordan, M., & Jones, R. (1999). A new test of social sensitivity: Detection of faux pas in normal children and children with Asperger syndrome. *Journal of Autism and Developmental Disorders*, 29, 407-418.
- Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, 19(3), 354-369.
- Bowler, D. M. (1992). "Theory of Mind" in Asperger's Syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 33(5), 877-893.
- Happé, F.G.E. (1994). An advanced test of theory of mind: Understanding story characters' thoughts and feelings by able autistic, mentally handicapped and normal children and adults. *Journal of Autism and Developmental Disorders*, 24, 1-24.
- 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.
- Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1430), 345-360.
- Lind, S. E. & Bowler, D. M. (2009). Language and theory of mind in autism spectrum disorder: The relationship between complement syntax and false belief task performance. *Journal of Autism and Developmental Disorders*, 39(6), pp. 929-937.
- Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., & Bishop, S. (2012). Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1-4. Torrance, CA: Western Psychological Services.
- Loth, E., Gómez, J. C., & Happé, F. (2008). Event schemas in autism spectrum disorders: the role of theory of mind and weak central coherence. *Journal of Autism and Developmental Disorders*, 38(3), 449-463.
- Malle, B. F., & Knobe, J. (1997). The Folk Concept of Intentionality. *Journal of Experimental Social Psychology*, 33(2), 101-121.
- Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences*, 108(7), 2688-2692.
- Zalla, T., Sav, A.-M., Stopin, A., Ahade, S., & Leboyer, M. (2009). Faux pas detection and intentional action in Asperger Syndrome. A replication on a French sample. *Journal of Autism and Developmental Disorders*, 39(2), 373-382.