



Chevallier, C; Noveck, I; Happe, F; Wilson, D; (2011) What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia*, 49 (3) 507 - 517.
[10.1016/j.neuropsychologia.2010.11.042](https://doi.org/10.1016/j.neuropsychologia.2010.11.042). Downloaded from UCL Discovery:
<http://discovery.ucl.ac.uk/1331364/>.

ARTICLE

What's in a voice? Prosody as a test case for the Theory of Mind account of autism

Coralie Chevallier^{1,*}, Ira Noveck², Francesca Happé³, Deirdre Wilson⁴

¹ *Laboratoire Langage, Cerveau et Cognition, CNRS - Université de Lyon; Institute of Psychiatry, King's College London*

² *Laboratoire Langage, Cerveau et Cognition, CNRS - Université de Lyon*

³ *Institute of Psychiatry, King's College London*

⁴ *Department of Linguistics, University College London and CSMN, University of Oslo*

* Corresponding author: Coralie Chevallier, Institut des Sciences Cognitives 67, Boulevard Pinel, 69675 Bron, France; E-mail: coralie.chevallier@gmail.com.

Abstract

The human voice conveys a variety of information about people's feelings, emotions and mental states. Some of this information relies on sophisticated Theory of Mind (ToM) skills, while others are more simple and do not require ToM. This variety provides an interesting test case for the ToM account of autism, which would predict greater impairment as ToM requirements increase. In this paper, we draw on psychological and pragmatic theories to classify vocal cues according to the amount of mindreading required to identify them. Children with a high functioning Autism Spectrum Disorder and matched controls were tested in three experiments where the speakers' state had to be extracted from their vocalizations. Although our results confirm that people with autism have subtle difficulties dealing with vocal cues, they show a pattern of performance that is inconsistent with the view that atypical recognition of vocal cues is caused by impaired ToM.

Keywords: Autism Spectrum Disorders; Emotions; Voice; Pragmatics; Theory of Mind

1. Introduction

Impaired Theory of Mind (ToM) has been described as one of the core deficits behind the communicative and social impairments in Autism Spectrum Disorders (ASD). More specifically, by preventing access to the full range of mental states and efficient mindreading, a deficit in ToM would lead to abnormalities in social development, in communicative development, in empathy and in imitation, all of which require taking other people's perspective (Baron-Cohen, 1995, 2000). Deficits in ToM have been evidenced using a variety of techniques and tasks among which first- (Baron-Cohen, Leslie, & Frith, 1985) and second-order false belief tasks (Bowler, 1992; Ozonoff, Pennington, & Rogers, 1991; Ozonoff, Rogers, & Pennington, 1991) as well as other tests designed for older and higher functioning individuals who typically pass standard false-belief tasks (Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999; Happé, 1994). Recent tests of ToM have concentrated on participants' ability to recognise mental or emotional states from information conveyed in the eye region or voice (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Baron-Cohen, Wheelwright, & Jolliffe, 1997; Golan, Baron-Cohen, Hill, & Rutherford, 2007; Rutherford, Baron-Cohen, & Wheelwright, 2002).

In particular, the understanding of feelings, emotions or mental states conveyed in the voice provides an interesting test case for the ToM account of autism. Indeed, these types of vocal cues are extremely diverse (for a review, see McCann & Peppé, 2003) and call up ToM to a varying degree, ranging from simple physical states –such as calm or excited– that do not require any mindreading, to social emotions –such as embarrassment– that require first-order ToM, and even more complex cues –like irony– requiring a full fledged 2nd-order ToM. This variety provides a good test case for the ToM account of autism. Indeed, according to this theory, participants with an ASD should be more impaired as ToM requirements increase.

So far, however, this factor has not been controlled for and this might account for mixed results in the existing literature on vocal emotion recognition in autism. On the one hand, a host of studies highlight difficulties in detecting vocal cues to irony and sarcasm (Wang, Dapretto, Hariri, Sigman, & Brookheimer, 2001; Wang, Lee, Sigman, & Dapretto, 2006, 2007), and in matching vocally expressed affects with a static or dynamic facial expression (Hobson, Ouston, & Lee, 1988; Loveland, et al., 1995) or with line drawings of body postures and facial expressions (Hobson, 1986). On the other hand, other studies show no deficits in recognising facial emotions (Buitelaar, Van der Wees, Swaab-Barneveld, & Van der Gaag, 1999; Grossman, Klin, Carter, & Volkmar, 2000), vocal emotions (Boucher, Lewis, & Collis, 2000), or in matching faces and voices in video segments (Loveland, et al., 1997). Within a single test, it was also found that some items trigger important differences between controls and adults with ASD, whilst others are not discriminant (Golan, Baron-Cohen, & Hill, 2006; Golan, et al., 2007; Kleinman, Marciano, & Ault, 2001; Rutherford, et al., 2002).

In this paper, we draw on psychological research and pragmatic theory to distinguish various categories of vocal cues as a function of the amount of mindreading required to identify them. This analysis allows us to arrange vocal cues on a single continuum that can adequately test the hypothesis according to which deficits in reading the mind in the voice are caused by a ToM deficit.

1.1 Mindreading, pragmatics and emotion recognition.

Here we aim to organize vocal cues that are currently studied in the experimental literature according to the amount of mindreading required to identify them. We begin with the recognition of *manners of speech*, which relies on the mere identification of external cues. For example, recognising that someone is stuttering or shouting depends on the detection of specific acoustic cues (atypical speech rate or increased intensity, respectively). Though stuttering may be symptomatic of an internal state of stress, and shouting of an internal state of anger, it is nonetheless possible to detect a stutter or a shout without possessing the concept of stress or anger.

Vocal cues to *physical states* – our second category – such as tiredness or drunkenness, by contrast, provide evidence for specific *internal* states. Recognising that someone sounds tired thus requires the ability to retrieve specific acoustic information (lower pitch and amplitude) and to link this to the appropriate internal physical state (e.g., the speaker is tired). Note that these first two categories both relate to non-mental states and, as such, should be identifiable in the absence of any capacity to mindread; they differ in that vocal cues in the second category, but not the first, are linked to the recognition of an internal (physical) state.

By contrast, our third through fifth categories – basic emotions, social emotions and certain types of speaker's attitude – all relate to internal *mental* states. These three categories are further distinguished in terms of the different orders of *metarepresentational* ability they require. A metarepresentational ability is an ability to think about representations with a conceptual content - one's own thoughts and those of others, for instance (Noh, 2001; Sperber, 2000; Wilson, 2000). A prime example of a metarepresentational ability is ToM. As

we will show, some mental states (basic emotions) are automatically detected independently of ToM, some (social emotions) depend on first-order ToM (and hence require a first-order metarepresentational ability), and others (expressions of speakers' attitudes to thoughts attributed to others) require second-order ToM.

The recognition of *basic emotions* (such as fear, anger, sadness, happiness and disgust), our third category, does not necessarily require any metarepresentational ability. To identify fear or anger, it is not mandatory to represent what the experiencer is thinking, but merely to relate external cues to the corresponding mental states (Ekman, Sorenson, & Friesen, 1969). In line with this claim, it has been demonstrated that non-human mammals who are clearly unable to represent the contents of others' thoughts can nonetheless recognise basic emotions (Tate, Fischer, Leigh, & Kendrick, 2006). Furthermore, infants become sensitive to these emotions well before they can pass any sort of ToM test. By about 4 to 6 months of age, they come to distinguish basic emotions and respond differentially to faces showing different emotions (e.g., neutral, happy, sad) (Izard, 1994; Serrano, Iglesias, & Loeches, 1992). By about 9 to 12 months, they even start seeking out emotional information from other people's faces when they feel unsure about the valence of a stimulus (Moore & Corkum, 1994).

Social emotions, our fourth category, have been described as more "complex" than basic emotions, and are indeed acquired later in development (Tracy & Robins, 2004). Crucially for our point, recent work indicates that recognition of social, but not basic, emotions is linked to ToM, both in typical development and in ASD (Heerey, Keltner, & Capps, 2003). There is also evidence that basic and social emotions activate different neural networks, with the latter prompting more activity in areas implicated in ToM (Moll, et al., 2002; Takahashi, et al., 2008). Finally, whereas basic emotions spring from situations, memories, or perceptions that have immediate personal relevance, social emotions are linked to other people's point of view (Haidt, 2003; Tracy & Robins, 2004) and experiencing them thus requires some degree of perspective taking, which is itself intimately linked to 1st order ToM. Thus, both the experience of social emotions and the ability to recognise them in others requires 1st-order ToM. This feature could also be related to the fact that social emotions involve genuinely propositional attitudes, i.e. attitudes to mentally represented states of affairs, and are thus necessarily metarepresentational in nature. One cannot simply be ashamed or embarrassed: one has to be ashamed *of* some state of affairs that one has in mind, or embarrassed *about* some mentally represented state of affairs. By contrast, basic emotions do not necessarily have this characteristic: One can just feel happy or sad without being happy or sad *about* some mentally represented state of affairs.

Whereas first-order ToM makes it possible to attribute to others beliefs that may differ from one's own, second-order ToM makes it possible to attribute beliefs about the beliefs of others that may differ from one's own (Wilson, 2000), and is crucial to our fifth, and last, category. Understanding an utterance that expresses the speaker's attitude to a thought attributed to someone else should therefore require an extra layer of metarepresentation. To illustrate, consider the utterances in (1) and (2). To understand (1), the hearer needs to attribute to Mary the thought that it will rain tomorrow: that is, to represent Mary's thought. To understand (2), the hearer needs to attribute to Mary a thought about John's thoughts: that is, to represent a thought about another thought.

- (1) Mary believes that it will rain tomorrow.
- (2) Mary thinks that John believes that it will rain tomorrow.

It has been argued that this higher order ToM is needed to understand irony and other related phenomena (Carston, 2002; Sperber & Wilson, 1986/1995). To illustrate, consider example (3b), uttered in the deadpan tone of voice typical of irony:

- (3) a. Bill (before the lecture): This will be a really good lecture.
b. Jane (later, sarcastically): That was a great lecture.

Sperber and Wilson (Sperber & Wilson, 1981, 1986/1995) argue that in uttering (3b), Jane is metarepresenting a thought she attributes to Bill (i.e. the thought that the lecture would be great), and expressing her own sceptical, mocking or contemptuous attitude towards it. Recognising Jane's intended meaning therefore requires the ability to understand that she is expressing a *thought* (or attitude) about an attributed *thought*. Experimental studies have indeed confirmed that performance on 2nd order ToM tasks is a good predictor of irony comprehension, both in typical and atypical development and in children with ASD (e.g., Adachi, et al., 2004; Happé, 1993; Langdon, Davies, & Coltheart, 2002).

Irony is only one of many types of case where recognising the speaker's intended meaning requires higher-order metarepresentational skills. Suppose, for instance, that Peter and Mary have gone out to see a film and, as they come out of the cinema, the following exchange occurs (example taken from Wilson, 2000):

- (4) Peter: That was a fantastic film.
Mary: [puzzled] Fantastic?

In (4), Mary echoes Peter's utterance not in order to agree or disagree with it, but in order to indicate that she is wondering about it. She thus expresses an attitude towards a thought she attributes to Peter. Just like ironical statements, then, so-called echoic questions require second-order ToM (Blakemore, 1994; Noh, 1998; Wilson, 2000). Another example of such dissociations between content and speaker's attitude can be found in some instances of echoic negation, as in the following exchange (5):

- (5) Peter: You look happy!
Mary: I'm not *happy*, I'm *ecstatic*.

These utterances involving metalinguistic negations are typically pronounced with a specific contradiction intonation contour (involving a final rise within the negative clause) and contrastive stress on the offending item (here, "happy") and on its correction (here, "ecstatic") in the second clause (Carston, 1996, 2002; Horn, 1985; Iwata, 1998). Here, Mary indicates her attitude (here, one of rejection) to a thought she attributes to Peter (i.e., the thought that Mary is happy).

Ironical utterances, echoic questions and metalinguistic negations all exhibit the same feature: an utterance (or part of an utterance) is used not to express the speaker's own thought, but to metarepresent a thought she attributes to someone else (or to herself at another time), and to convey her own actual attitude to the content of that thought (Noh, 2001; Wilson, 2000, 2006, in press). In each case, two orders of metarepresentational ability are required to recognise the speaker's meaning.

As this analysis makes clear, vocal cues provide several different types of evidence. Some reveal no more than the manner in which speech is produced, while others reveal underlying physical or mental states. Moreover, the cues that provide evidence of underlying mental states require different orders of metarepresentational or mindreading ability: some (basic emotions) require no mindreading ability at all, others (social emotions) require first-order mindreading ability, while still others (2nd order attitudes) require second-order mindreading ability. We might therefore expect these different categories of vocal cue to be processed in different ways.

1.2 Goals of the present paper

In this paper, we aim to test whether participants with an ASD are more impaired in processing vocal cues that rely on ToM. We present an experimental procedure designed so that: All categories of vocal cues can be compared, both reaction times and accuracy rates are measured, no complicated vocabulary is used, a satisfactory number of items is included, and content effects are overridden. Experiment 1 assesses the ability to recognise manners of speech, physical states, basic emotions, social emotions and 2nd order mental states solely from vocal cues, and compares performances in ASD and TD participants. Experiments 2 and 3 then investigate the potential use of compensatory strategies in the ASD group. Following the ToM account, participants in the ASD group should have lower performance rates and slower reaction times in conditions requiring mentalising (social emotions and 2nd order condition) but control-like performances for non-mentalistic items (manners of speech, physical states, and basic emotions).

2. Experiment 1

Experiment 1 assesses the ability to take a variety of vocal cues into account in order to retrieve information about the speaker's physical or mental state. Sentences with a neutral content and a marked prosodic contour were presented to the participant, who then had to pick the follow-up sentence – out of two – which best describes the way the speaker feels. For example, the item “Ben hears a big noise from his neighbours’ house. He says: *What is that noise?*” was presented, followed by two possible choices: “Ben is scared. There might be a burglar in his neighbours’ house!” and “Ben is angry. He doesn’t like it when his neighbours are too noisy.” To ensure that there was no content effect, each sentence was associated with two distinct intonation types (appearing in different lists). In our example, the target sentence (in italics) could be uttered either with a stutter or in a singing voice. Reaction times and accuracy rates were measured in five conditions: i) Manners of speech condition, ii) Physical states condition, iii) Basic emotions condition, iv) Social emotions condition, and v) the 2nd order condition.

2.1 Methods

2.1.1 Participants

Thirty-four male adolescents (17 with ASD and 17 Typically Developing, henceforth TD) took part in Experiment 1. The pupils with ASD were at special education schools in England which require formal diagnosis of an ASD according to standard clinical criteria (APA, 1994). The diagnostic information was gathered from school files of documented medical diagnoses made by a clinical psychologist and/or psychiatrist. The controls were seen in a regular school. TD and ASD participants all spoke English at home, and none had any significant hearing loss, visual impairment, or major physical disability. The control group was matched on chronological age (ASD-Mean = 13;8, TD-Mean = 14;2) and verbal mental age (Standardised BPVS score: ASD-Mean = 106, TD-Mean = 100, see Table 1 for detailed information).

INSERT TABLE 1 ABOUT HERE

We also matched our samples on their ability to discriminate pitch, duration and intensity using Dorothy Bishop's Dinos tasks (for previous studies using the Dinos task, see, e.g. Jones, et al., 2009; Sutcliffe & Bishop, 2005). Two dinosaurs each make a sound separated by a 500ms interval in the intensity and duration tasks and by a 480ms interval in the frequency task. The child then has to decide which dinosaur is making the longest, loudest or highest sound (depending on whether she is completing the duration, intensity or frequency task, respectively). Correct responses in this task are reinforced with a small icon on the screen and a cheerful noise, and wrong answers with a cross and a sigh noise. The next trial starts after a 500ms interval. All three tasks are based on a “more virulent” PEST procedure (Findlay, 1978), which adaptively alters the gap separating the two sounds. Initially, the participant has to make very easy discriminations, and difficulty is gradually

increased until an error is made. When an error is made, the discrimination is made easier. The task is stopped after 6 reversals have occurred or a maximum of 40 trials has been completed. The PEST procedure is set to converge on the 75% correct point and the threshold is taken as the average target across the last four reversals in the track. Minimum discrimination thresholds for the three Dinos tasks are shown in Table 2. Note that low thresholds are indicative of optimal performance.

INSERT TABLE 2 ABOUT HERE

2.1.2 Material and design

Design. Each target sentence was followed by two options and preceded by a sentence setting up a little context designed to make the sentence as natural as possible. Each item was associated with two different prosodic cues and two different possible answers. Each item could be associated with either answer, depending on the prosodic cue that was used. Two lists were set up, each including all 32 items pronounced in one of their prosodic form only. Depending on the prosodic cue, one option was correct and the other incorrect. This design allowed us to ensure that the participants' responses could not be influenced by the semantic content of the item and that they had to rely solely on vocal cues. Thirty-two items were included in the study. The 2nd order condition included 8 items, and all other conditions included 6 items (for a detailed list of stimuli, see Appendix A).

Auditory stimuli. The stimuli were recorded in an anechoic chamber at University College London with the help of a professional acoustician. The speaker was a native male speaker of Southern standard British English, trained to record auditory stimuli. He sat in an armchair equipped with a headrest ensuring that the distance between his mouth and the microphone remained constant. The microphone (Bruel & Kjaer 2231 Sound Level Meter fitted with a Type 4165 Microphone) was linked up to a Sony DAT reader connected to a PC. The recordings were made in a mono format, using a 44.1 kHz sampling rate. The items to be read were presented on a suspended computer screen using ProRec version 1.0© (Huckvale, 2003). The wave files were then segmented using the Speech Filing System© (Huckvale, 2004) and a 100 ms silence was inserted immediately before and after the sound signal.

Testing. The experiment was presented using a laptop and the sounds were played through Sennheiser headphones that were calibrated for consistency of dB before use. None of the children had problems agreeing to wear the headphones and all of them were comfortable with computers.

2.1.3 Procedure

Written parental consent was obtained prior to the testing phase and children were then asked for oral assent. Pupils were seen individually at school, in a quiet room. The experiment started with the following instructions, which were presented on the screen and read out loud to the participant:

You are going to hear a man called Ben talking and you will need to pick the sentence which describes best the way he feels. For example, if Ben sounds sad, you will need to pick the description which says that Ben is sad. You will need to choose between two sentences: one description will be written in red, and one description will be written in blue. If you think the best description is the red one, press the RED button, if you think the best description is the blue one, press the BLUE button.

The instructions were followed by a three trial training phase after which the participant could ask questions. The experimental phase then started. When participants were halfway through the task, the message "You're half way through!" was displayed so that they had the

opportunity to take a break. Each trial started with a 1000ms “Listen carefully” screen followed by an auditory stimulus. The participant then had to answer using one of two response keys (E and P counterbalanced) and the next trial started 1000ms later. The trials were presented in a random order.

2.2 Data analysis for Experiments 1 to 3

The data was analysed using Statistica 7.1. For all reaction time analysis, a log transformation was carried out beforehand to improve the conformity of the data to the standard assumptions of ANOVA (e.g., Howell, 1997). Reaction times of more than three standard deviations from the mean were considered outliers and were excluded from both the reaction time and the choice proportion analysis. Moreover, only correct responses were retained in the reaction time analysis. Effect sizes were also calculated using r^2 . Following Cohen (1988), an r^2 above .010 reflects a small effect, an r^2 above .059 reflects a medium effect, and an r^2 above .138 reflects a large effect. Mixed repeated measures ANOVA with two factors were conducted: the within-subject factors “Prosodic-Cue” (5: speech manners, physical states, basic emotions, social emotions and 2nd order mental states) and the between subject factor “Group” (2: TD, ASD). The dependent variable was either the rate of correct responses or reaction times. All p -values assume a two-tailed test.

2.3 Results

Interestingly, the pattern of performances in the various conditions reflected the one predicted by our theoretical distinctions: Manners of speech were easiest to detect (Mean = 91.1), followed by the other non-mentalistic category (Physical states Mean = 83.2). The three mental categories became harder as the number of metarepresentations needed to retrieve the information increased. For basic emotions, percentage of correct responses went down to 76.8; social emotions, which require first order ToM, came next (Mean = 74.8), followed by 2nd order stimuli (Mean = 70.1), the hardest category (see Figure 1). A Spearman rank order correlation confirms this gradual decrease of performance as the complexity of mentalising increases, $r = -.34$; $p < .0001$.

INSERT FIGURE 1 ABOUT HERE

The ANOVA reveals a main effect of Prosodic-Cue, $F(4,128) = 9.5$, $p < .001$; $r^2 = .22$. *Post hoc* Tukey tests indicate that the Manners of speech condition was easier than all the others, all $ps < .05$. The Physical states condition also differed from all conditions, all $ps < .05$, except for the basic emotions condition for which there was only a trend, $p = .09$. However, there was no significant main effect of the group, $F(1,32) = 0.13$, $p = .72$; $r^2 = .004$, and no Prosodic-Cue X Group interaction, $F(4,128) = 0.03$, $p = .99$; $r^2 = .001$. We then turned to a more detailed analysis in order to determine whether ASD participants and TD participants not only had similar global performances, but also similar patterns of successes and difficulties. With this goal in mind, we compared ASD and TD performances for each item and checked whether patterns of performances correlated. A strong correlation was found between ASD and TD participants' performances, $r = 0.66$; $p < .0001$; $r^2 = 0.44$. Strong correlations between the two groups were also found when each condition was considered (Manners of speech: $r = 0.86$; $p < .001$; $r^2 = 0.74$; Basic Emotions : $r = 0.64$; $p < .05$; $r^2 = 0.41$; Social Emotions: $r = 0.61$; $p < .05$; $r^2 = 0.37$; 2nd Order Mental States: $r = 0.64$; $p < .01$; $r^2 = 0.41$); with the exception of the Physical States condition ($r = 0.39$; $p = .22$; $r^2 = 0.15$). This indicates that items that were difficult for one group were also difficult for the other group; and, conversely, that items that were easy for one group were also easy for the other group.

Furthermore, the analysis of reaction times reveals a pattern similar to that observed for accuracy rates, with a main effect of Prosodic-Cue, $F(4,128) = 5.47$, $p < .001$; $r^2 = .15$, no main effect of group, $F(1,32) = 0.04$, $p = .85$; $r^2 = .001$, and no Prosodic-Cue X Group

interaction, $F(4,128) = 0.64$, $p = .63$; $r^2 = .019$, which suggests that both groups of participants processed the stimuli at similar speeds.

2.4 Discussion

The data collected in Experiment 1 highlight the relevance of the theoretical distinction made in the introduction. Indeed, the pattern of accuracy rates in response to the various categories of vocal cues parallels distinctions based on prerequisites for mindreading. More surprisingly, though, ASD participants were unimpaired in all conditions. First, ASD and TD participants were as accurate in processing all sorts of prosodic cues, including those requiring greater metarepresentational abilities. Second, the pattern of performance covaried in both groups, so that items which triggered low (or high) performances in one group also triggered low (or high) performances in the other group. Third, ASD participants were as fast as TD participants in making their judgments. Overall, these results appear to contradict the idea that individuals with an ASD have a specific impairment in recognising ToM related emotions.

One way to account for these data is to argue that, whilst failures at pragmatic tasks indicate pragmatic deficits, passes do not necessarily reflect underlying competence and might reflect the use of compensatory strategies. The rationale for this line of argument is that people with ASD, being particularly verbally able, can *explicitly* reason about the mental states of others, whilst being unable to mentalise in a more intuitive fashion (Fisher, Happé, & Dunn, 2005; Happé, 1995). This line of reasoning also applies to our data: although performance is very similar across groups, one cannot exclude the possibility that ASD and TD participants resort to different underlying cognitive mechanisms to make their judgments, and rely on compensatory strategies which lead them to be as accurate and efficient as the controls. This claim has several consequences. First, if compensatory mechanisms are used, it should be possible to disrupt them by making the task more demanding. For instance, individuals with an ASD who do pass 2nd order ToM tasks often remain impaired in more subtle tests (e.g., Baron-Cohen, et al., 2001; Happé, 1994). Second, if such compensatory mechanisms are rooted in high verbal skills, performance in pragmatic tasks should correlate with measures of verbal intelligence (Fisher, et al., 2005; Happé, 1995). Finally, compensatory strategies are likely to be less efficient than genuine competence and should thus give rise to slower reaction times. Such claims have rarely been directly tested empirically and are addressed in Experiment 2.

3. Experiment 2

Experiment 2 assesses the ability to recognise a variety of vocal cues whilst involved in a dual task. Participants' first task was to decide – as fast as they could – whether or not they had heard the sound “ing” in the spoken stimulus. The interfering task was presented as the primary task and the emotional task was presented as the secondary task. If ASD participants rely on compensatory strategies to make up for an impaired ToM, they should be slower in conditions where mindreading is essential (Social Emotions and 2nd Order Mental States); following the disruption caused by the dual task, their accuracy rates in those conditions should drop; finally, their performance should correlate with verbal intelligence.

3.1 Methods

3.1.1 Participants

Forty male adolescents (20 with AS and 20 TD) took part in Experiment 2. Inclusion and matching criteria were identical to those used in Experiments 1 (see Tables 1 and 2 for detailed information).

3.1.2 Materials

The same auditory stimuli and materials were used in Experiments 1 and 2. However, given that Experiment 2 was more demanding, we decided to reduce the number of trials and

excluded half of the items (see Appendix A: starred items appear in all three experiments). Items which had elicited the poorest rates of performance in the TD group were excluded.

3.1.3 Procedure

The experiment started with the instructions for the Detection task:

You're going to listen to two men talking. Your job is to say whether you heard them pronounce the sound "ing". Answer as fast as you can! Ready?

This was followed by a four trial training phase. After this phase, the second task (the Emotion task) was introduced in the same way as in Experiment 1. The rest of the procedure was identical to the one used in Experiment 1. Throughout the whole procedure, participants first had to say whether or not they had heard the sound "ing" by pressing "yes" or "no" on the keyboard. They received feedback on their speed and accuracy immediately after providing their answer. The feedback screen remained for 1 second and was followed by the emotion question.

3.2.1 Results Task 1: Detecting the sound "ing"

A one way ANOVA comparing rates of correct answers in the ASD (Mean = 89 %) and TD (Mean = 93 %) groups indicates no significant difference, $F(1,38) = 2.46$, $p = .13$; $r^2 = .06$. The same analysis for reaction times also reveals no group differences (ASD Mean = 4225 ms, TD Mean = 3711 ms, $F(1,38) = 0.36$, $p = .55$; $r^2 = .009$).

3.2.2 Results and discussion Task 2: External and mental state recognition

As in Experiment 1, manners of speech were easiest to detect (Mean = 96.3), followed by the other categories requiring no mindreading (Physical states Mean = 83.8; Basic Emotions Mean = 88.4). The categories requiring first or second order ToM were the hardest (Social Emotions Mean = 75.0; 2nd Order Mental States Mean = 77.7) (see Figure 2). As in Experiment 1, a Spearman rank order correlation confirms that there is a gradual decrease of performance as the complexity of mentalising increases, $r = -.37$; $p < .0001$.

INSERT FIGURE 2 ABOUT HERE

The ANOVA revealed a main effect of Prosodic-Cue, $F(4,152) = 5.7$, $p < .001$; $r^2 = .13$) but no significant main effect of the group, $F(1,38) = 0.08$, $p = .78$; $r^2 = .002$, and no Prosodic-Cue X Group interaction, $F(4,152) = 0.11$, $p = .98$; $r^2 = .003$. As in Experiment 1, the performances of ASD and TD participants correlate, $r = 0.38$; $p < .05$; $r^2 = 0.14$, which indicates similar patterns of successes and difficulties in both groups.

As in Experiment 1, the analysis of reaction times reveals a main effect of Prosodic-Cue, $F(4,144) = 2.86$, $p < .05$; $r^2 = .07$. There was also no Group X Prosodic Cue interaction, $F(4,144) = 0.69$, $p = .60$; $r^2 = .018$ and no main effect of group, $F(1,36) = 0.00$, $p < .98$; $r^2 = .0001$, which indicates that ASD participants provided their answers as quickly as TD participants.

Finally, we found that global performance did not correlate with BPVS scores in the ASD group, $r = 0.24$; $p = .30$; $r^2 = .06$, or in the TD group, $r = -0.12$; $p = .64$; $r^2 = .03$. The same pattern was found for reaction times, ASD group: $r = 0.02$; $p = .94$; $r^2 = .0003$; TD group: $r = 0.31$; $p = .19$; $r^2 = .10$.

3.3 Discussion

In Experiment 2, we tested participants' ability to identify vocal cues whilst they were engaged in a secondary task. By doing so, we aimed to uncover potential compensatory strategies which may have masked differences in Experiment 1. The use of compensatory strategies has several testable consequences. First, an increase in task difficulty should lead

to worse performances in conditions requiring mindreading. Second, performance should correlate with measures of verbal intelligence. Finally, since compensatory strategies are, by nature, less efficient than the actual cognitive process they replace, participants who resorted to compensatory strategies (here, ASD participants) should be slower in conditions requiring ToM abilities.

None of these predictions was supported. In spite of the increased cognitive demands imposed by the dual task, ASD participants were as accurate and fast as the controls; they had the same pattern of strengths and deficits; and finally, performances did not correlate with measures of verbal intelligence in either group. The two groups also had similar performances in the interfering task, which implies that the lack of difference in the emotion task is unlikely to be due to ASD participants being less attentive in the interfering task. Together with Experiment 1, these results suggest that the participants with ASD included in our study are capable of genuinely “reading the mind in the voice”. An alternative possibility, though, is that the task was not challenging enough to pinpoint a subtle, but existing, deficit. In Experiment 2, everything was indeed done to ensure that the dual task was easy and engaging for the children. Observation during the experimental session and also debriefing confirmed that participants in both groups enjoyed the task and got excited about using the feedback screen to try and beat their own speed record. Because the participants included in this study were all high functioning, it is possible that their emotion recognition deficit is restricted to challenging situations. In Experiment 3, we address this possibility.

4. Experiment 3

Experiment 3 assesses the ability to detect and interpret a variety of vocal cues whilst involved in a highly demanding dual task. Participants were asked to concentrate on the number of times they heard the letter T in the utterance whilst also having to monitor the speaker’s emotional state. The interfering task (counting Ts) was presented as the primary task and the emotional task was presented as the secondary task. The predictions were identical to those made in Experiment 2: If ASD participants rely on compensatory strategies to make up for an impaired ToM, they should be slower in conditions where mindreading is essential (Social Emotions and 2nd Order Mental States); following the disruption caused by the dual task, their accuracy rates in those conditions should drop; and finally, their performance should correlate with verbal intelligence.

4.1 Methods

4.1.1 Participants

Thirty-two male adolescents (16 with ASD and 16 TD) took part in Experiment 3. Inclusion and matching criteria were identical to those used in Experiments 1 and 2 (see Tables 1 and 2 for detailed information).

4.1.2 Materials

The same material as presented in Experiment 2 was used.

4.1.3 Procedure

The experiment started with the instructions for the Counting task:

You’re going to listen to two men talking and you will have to count the number of times you hear the letter “T” in what they say. You need to pay close attention as the two men in the game sometimes talk quite fast.

This was followed by a four trial training phase. After this phase, the second task (the Emotion task) was introduced:

Your second job is to decide how Ben feels. Do you remember how Ben felt in what we just listened to? Did Ben sound tired or happy? For the next examples, you will

see two sentences describing how Ben could possibly feel and you will need to choose the best one. For example, there could be a sentence saying that Ben is tired because he worked too hard and another sentence saying that Ben is happy because he got a nice present. One sentence will be written in red and one sentence will be written in blue. If you think the best description is the red one, press the RED button, if you think the best description is the blue one, press the BLUE button. Now we are going to practice a little more and then the real game will start. Remember to count the Ts in what you hear.

These instructions were followed by a second training phase including two trials. The experimental phase then started, and followed the same procedure as in Experiments 1 and 2. Throughout the whole procedure, participants were asked to say the number of Ts out loud and the experimenter recorded her answer. The emotion question then appeared on the computer and the participant provided her answer by pressing the appropriate key.

4.2.1 Results Task 1: Counting Ts

A mixed repeated measures ANOVA was performed with the within subjects factor "Actual number of Ts" (2, 4, 5, 6, 7, 8) and the between subjects factor Group (TD, ASD). The dependent variable was the number of Ts that were detected by the participants. The analysis revealed no main effect of Group, $F(1,30) = 0.50$, $p = .49$; $r^2 = .016$, and a main effect of the Actual number of Ts on participants' answers, $F(5,150) = 27.84$, $p < .00001$; $r^2 = .48$. This main effect indicates that the number of Ts detected by participants varied depending on the actual number of Ts present in the utterance (see Figure 3). However, the absence of a main effect of Group indicates that ASD and TD participants performed similarly. In other words, it appears that both groups were equally involved in the Counting Ts task.

INSERT FIGURE 3 ABOUT HERE

4.2.2 Results Task 2: External and mental states recognition

As observed before, manners of speech were easiest to detect (Mean = 95.3), followed by the other categories requiring no mindreading (Physical states Mean = 85.9; Basic Emotions Mean = 90.0). The categories requiring first or second order ToM were hardest (Social Emotions Mean = 76.6; 2nd Order Mental States Mean = 76.6) (see Figure 4). As in Experiments 1 and 2, a Spearman rank order correlation confirms the existence of a gradual decrease of performance as the complexity of mentalising increases, $r = -.33$; $p < .0001$.

INSERT FIGURE 4 ABOUT HERE

The ANOVA reveals a main effect of Prosodic-Cue, $F(4,120) = 6.2$, $p < .001$; $r^2 = .17$, but no significant main effect of the group, $F(1,30) = 0.02$, $p = .90$; $r^2 = .0005$, and no Prosodic-Cue X Group interaction, $F(4,120) = 0.36$, $p = .84$; $r^2 = .012$. As in Experiments 1 and 2, a strong correlation was found between ASD and TD participants' performance, $r = 0.67$; $p < .0001$; $r^2 = 0.44$, which indicates similar patterns of successes and difficulties.

The analysis of reaction times reveals a main effect of Prosodic-Cue, $F(4,116) = 5.74$, $p < .001$; $r^2 = .17$, and, in contrast to what had been found previously, a main effect of group, $F(1,29) = 9.89$, $p < .01$; $r^2 = .25$, due to ASD participants being slower than TD participants (ASD Mean = 7074 ms; TD Mean = 5081 ms; Tukey test: $p < .005$). There was no Group X Prosodic Cue interaction, $F(4,116) = 0.62$, $p = .65$; $r^2 = .020$, which suggests that ASD participants were slower overall but were not especially impaired in the ToM conditions.

Finally, we found that global performance and reaction times did not correlate with BPVS scores in the ASD group, $r = 0.25$; $p = .35$; $r^2 = .06$ and $r = 0.18$; $p = .50$; $r^2 = .03$,

respectively. In the TD group, this correlation was marginally significant for both measures, $r = 0.48$; $p = .07$; $r^2 = .23$ for accuracy rates, and $r = 0.47$; $p = .07$; $r^2 = .22$ for reaction times.

4.3 Discussion

In Experiment 3, we tested participants' ability to identify vocal cues whilst they were engaged in a highly demanding secondary task. In this new task, we replicated the results found in the previous two experiments: ASD participants were as accurate as the controls; they had the same pattern of strengths and deficits; finally, performance did not correlate with measures of verbal intelligence in either group. There was even a trend in the opposite direction, with a marginally significant correlation in the TD group only. The prediction regarding reaction times was also not verified, since ASD participants were not *specifically* slowed down in conditions requiring mindreading. Instead, ASD participants were slower in *all* conditions, which suggests that when placed under cognitive load, participants with ASD have difficulties identifying vocal cues *in general*, independently of underlying mindreading requirements.

5. General discussion

In this paper, we used prosody as a test case for the ToM account of autism. We argued that the various types of cues conveyed in the voice call up ToM to a varying degree, ranging from simple physical states, to social emotions or more complex cues that require a full fledged ToM. We presented a series of three experiments designed to assess whether participants with an ASD are more impaired in processing vocal cues that rely on ToM. With this goal in mind, we proposed a set of categories by distinguishing cues that require various levels of mentalising, and compared performance in ASD participants and TD participants. Contrary to the predictions of the ToM account, ASD participants were not specifically impaired in conditions requiring higher order mindreading skills. In Experiment 1, ASD and TD participants had similar accuracy rates and reaction times across all conditions, and they displayed the same pattern of strengths and difficulties. This was confirmed in Experiment 2 despite the increased demands imposed by the dual task. Finally, in Experiment 3, ASD participants showed no ToM-specific impairment in a highly demanding dual task. On the contrary, we observed that they were slower than TD participants in *all* conditions, which suggests that, when placed under high cognitive load, they have difficulties identifying vocal cues in general, independently of underlying mindreading requirements.

Taken together, this absence of a specific impairment in conditions related to ToM suggests that our sample of ASD participants is capable of genuinely reading the mind in the voice. However, the overall slow-down in reaction times observed in Experiment 3 indicates that they have difficulty making use of vocal cues in challenging situations. This is in line with numerous findings highlighting the gap between their performance in structured experimental tasks and in real life situations (for a review, see Klin, 2003). Real-life social situations are especially demanding because many crucial social cues are made available in parallel and need to be rapidly integrated in order to make sense of the situation. This might also explain some of the discrepancies between experimental results demonstrating no emotion recognition deficit (see e.g., Adolphs, 2001; Boucher, et al., 2000; Buitelaar, et al., 1999; Loveland, et al., 1997) and the actual experience of people with autism in their daily life. There is indeed an important difference between experimental settings where one is explicitly told to look out for emotional cues and complex social environments providing no guidance as to what should be attended to. In real life then, "the individual needs to go about defining a social task as such by paying attention to, and identifying, the relevant aspects of a social situation prior to having an opportunity to use their available social cognitive problem-solving skills" (Klin, 2003, p. 347). In sum, our own results confirm that people with autism have difficulties dealing with emotional cues in challenging contexts; yet they undermine the idea that impaired ToM is at the core of this deficit.

In line with this conclusion, previous studies on vocal cue recognition do not seem to highlight specific deficits in items linked to ToM. In the original Reading the mind in the voice task, for instance, Rutherford et al. (2002) found deficits in basic emotion items (e.g. *joyous* vs. *scared*) but control-like performances on some social emotion items (e.g. *disappointed* vs. *apologetic*) or speaker's attitudes (e.g. *sarcastic* vs. *indifferent*). Interestingly, when more foils were added in the revised version of the task (Golan, et al., 2007), performance in the ASD group became worse. But again, the worsening was not specific to ToM related emotions: indeed, performance remained identical for some social emotions (e.g. *worried* or *apologetic*) but worsened for some physical state items (e.g. *nervous*) or some basic emotion items (e.g. *terrified*). In another study, Golan et al. (Golan, et al., 2006) also point out that there was no difference between the autism and control groups in the recognition of empathy. Equally surprising was the lack of difference in recognizing mental states such as "appalled", which is a social emotion. As the authors mention, these results suggest that "mindblindness" is by no means total. Conversely, the claim that the emotion recognition deficit is not specific to ToM is in line with findings indicating impairments in basic emotion recognition (see e.g., Celani, Battacchi, & Arcidiacono, 1999; Deruelle, Rondan, Gepner, & Tardif, 2004; Hobson, 1986; Hobson, et al., 1988; Loveland, et al., 1995). As noted above, however, people with ASD are sometimes capable of displaying control-like performances, which suggests that the underlying competence to recognize emotions is there, but might be blocked by some other factor.

Children and adolescents with autism do indeed show various signs of emotional understanding: they recognize emotional expressions in certain contexts, they refer to complex emotions such as pride or embarrassment at the same rates as their typically developing peers, and they acknowledge that emotional states in others influence their behaviour (for a review, see Begeer, Koot, Rieffe, Meerum Terwogt, & Stegge, 2008). What particularly distinguishes them from matched controls is their lack of spontaneous bias towards seeking social cues and the peculiar nature of their comments on emotions, which have been described as idiosyncratic, underinformative, scripted or lacking pragmatics and references to social causes (for a review, see Begeer, et al., 2008). In a related fashion, results from a different study indicate that children with Asperger Syndrome are able to understand facial cues of emotion but are less likely to seek them in more demanding affective processing tasks (Grossman, et al., 2000).

The main deficit, then, might be one of diminished social orienting or diminished social motivation (Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998; Klin, 2003). People with autism indeed appear to be less predisposed to orient to salient social stimuli (perhaps because they fail to see their intrinsic rewarding value, Dawson, Webb, & McPartland, 2005; Dawson, Webb, Wijsman, et al., 2005) and might be less motivated to solve social problems (Klin, 2003). This hypothesis predicts that performance in emotion recognition tasks will be boosted if social orienting is enhanced by extrinsic factors. For instance, in a recent study on the neural correlates of irony comprehension in autism, Wang and collaborators (Wang, et al., 2007) compared neutral instructions ("Pay close attention") and explicit social instructions ("Pay close attention to the face and voice"). They demonstrated that activity in the medial prefrontal cortex, which is activated when TD participants interpret ironical utterances, increased in the ASD group in the explicit condition. A similar effect of explicit instructions was also recently found in an oddball task where participants heard both speech and non-speech sounds. In line with previous research (Ceponiene, et al., 2003), children with autism had atypical ERP (Event Related Potentials) profiles in response to speech sounds, but not to non-speech sounds. However, this difference disappeared when participants were *explicitly* required to pay attention to the sound stream.

Similarly, the processing of emotional cues in faces appears to be related to task demands. For instance, in a spontaneous photograph sorting task where two possible criteria – emotional and non emotional (e.g. the identity of the person in the photograph) – can be

used, individuals with (low functioning) autism often prefer non-emotional sorting criteria while TD participants spontaneously favour the emotional ones (Davies, Bishop, Manstead, & Tantam, 1994; Weeks & Hobson, 1987). However, this difference disappears when the emotional criterion is made relevant (i.e., "Which ones would be likely to give you a sweet?"; Begeer, Rieffe, Terwogt, & Stockmann, 2006). Finally, the participant's own intrinsic motivation to attend to social stimuli can also be influential. For instance, Kahana-Kalman and Goldman (Kahana-Kalman & Goldman, 2008) demonstrated that four-year-old children with ASD were better at matching facial and vocal expressions of emotion when these were portrayed by their mother, compared to an unfamiliar adult.

To conclude, our results, along with other past empirical findings, show a combination of competences and impairments among those with ASD which is inconsistent with the idea that atypical recognition of vocal cues is caused by impaired ToM. The relative ease with which children with ASD manage to process vocal cues also suggests that some aspects of emotional and pragmatic processing are spared in at least a subgroup of individuals on the autism spectrum. Future work will need to carefully characterise this subpopulation, the nature of their social deficit, and the scope of the emotional and pragmatic processes they can deal with. In particular, one limitation of the present study is the absence of confirmatory ASD diagnoses using the gold standard ADI-R and ADOS (Lord, et al., 2000). Apart from ensuring the validity of the diagnosis, including such clinical measures in the future would allow for a better characterisation of the subgroups who pass emotional and pragmatic tests.

At a conceptual level, the possibility that factors such as social attention and social motivation play an important role is in urgent need of further empirical evidence and should be systematically investigated. Apart from raising important theoretical issues, this topic also has crucial clinical implications. If the cognitive devices that enable vocal cue recognition are indeed spared in autism, educational strategies should then directly tackle the motivational or attentional deficits which prevent people with autism from appropriately resorting to them.

Acknowledgments

Many thanks to the children and staff in North Hill House (Frome, UK), Southlands (Lymington, UK), Henry Fanshaw School (Dronfield, UK), Chelmer Valley High School (Chelmsford, UK) and Haberdasher's Aske's Boys School (Herts, UK). We also wish to thank Dorothy Bishop for permission to use her Dino task programme, Jenny Thomson and Usha Goswami for providing a modified version of the task, and Catherine Jones for valuable advice on the programme. Many thanks, finally, to Steve Nevard for technical support and Tim Wharton for the recordings.

References

- Adachi, T., Koeda, T., Hirabayashi, S., Maeoka, Y., Shiota, M., Charles Wright, E., et al. (2004). The metaphor and sarcasm scenario test: a new instrument to help differentiate high functioning pervasive developmental disorder from attention deficit/hyperactivity disorder. *Brain and Development*, 26(5), 301-306.
- Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, 11(2), 231-239.
- Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge: MIT Press.
- Baron-Cohen, S. (2000). Theory of mind and autism: a 15-year review. In S. Baron-Cohen, H. Tager-Flusberg & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (pp. 3–21). Oxford: Oxford University Press.
- 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.
- Baron-Cohen, S., Wheelwright, S., Hill, 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. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(02), 241-251.
- 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.
- Begeer, S., Koot, H., Rieffe, C., Meerum Terwogt, M., & Stegge, H. (2008). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. *Developmental Review*, 28(3), 342-369.
- Begeer, S., Rieffe, C., Terwogt, M., & Stockmann, L. (2006). Attention to facial emotion expressions in children with autism. *Autism*, 10(1), 37.
- Blakemore, D. (1994). Echo questions: A pragmatic account. *Lingua*, 4, 197-211.
- Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice Processing Abilities in Children with Autism, Children with Specific Language Impairments, and Young Typically Developing Children. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41(07), 847-857.
- Buitelaar, J., Van der Wees, M., Swaab-Barneveld, H., & Van der Gaag, R. (1999). Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. *Development and psychopathology*, 11(1), 39-58.
- Carston, R. (1996). Metalinguistic negation and echoic use. *Journal of Pragmatics*, 25(3), 309-330.
- Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*: Blackwell Publishers.
- Celani, G., Battacchi, M., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders*, 29(1), 57-66.
- Ceponiene, R., Lepisto, T., Shestakova, A., Vanhala, R., Alku, P., Naatanen, R., et al. (2003). Speech-sound-selective auditory impairment in children with autism: They can perceive but do not attend. *Proceedings of the National Academy of Sciences*, 100(9), 5567-5572.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed. ed.). Hillsdale, NJ: Erlbaum.
- Davies, S., Bishop, D., Manstead, A., & Tantam, D. (1994). Face perception in children with autism and Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 35, 1033 - 1057.
- Dawson, G., Meltzoff, A., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with Autism Fail to Orient to Naturally Occurring Social Stimuli. *Journal of Autism and Developmental Disorders*, 28(6), 479-485.
- Dawson, G., Webb, S., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology*, 27(3), 403-424.
- Dawson, G., Webb, S., Wijsman, E., Schellenberg, G., Estes, A., Munson, J., et al. (2005). Neurocognitive and electrophysiological evidence of altered face processing in parents of children with autism: implications for a model of abnormal development of social brain circuitry in autism. *Development and psychopathology*, 17(03), 679-697.
- Deruelle, C., Rondan, C., Gepner, B., & Tardif, C. (2004). Spatial frequency and face processing in children with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, 34(2), 199-210.
- Ekman, P., Sorenson, E., & Friesen, W. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164(3875), 86.

- Findlay, J. M. (1978). Estimates on probability functions: A more virulent PEST. *Perception and Psychophysics*, 23(2), 181–185.
- Fisher, N., Happé, F., & Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry*, 46(4), 409-419.
- Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders*, 36(2), 169-183.
- Golan, O., Baron-Cohen, S., Hill, J., & Rutherford, M. (2007). The 'Reading the Mind in the Voice' Test-Revised: A Study of Complex Emotion Recognition in Adults with and Without Autism Spectrum Conditions. *Journal of Autism and Developmental Disorders*, 37(6), 1096-1106.
- Grossman, J., Klin, A., Carter, A., & Volkmar, F. (2000). Verbal bias in recognition of facial emotions in children with Asperger syndrome. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41(03), 369-379.
- Haidt, J. (2003). The moral emotions. In D. R. K. Scherer & H. Goldsmith (Eds.), *Oxford Handbook of affective sciences* (pp. 852-870). Oxford: Oxford University Press.
- Happé, F. (1993). Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition*, 48(2), 101-119.
- Happé, F. (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.
- Happé, F. (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.
- Heerey, E. A., Keltner, D., & Capps, L. M. (2003). Making sense of self-conscious emotion: Linking theory of mind and emotion in children with autism. *Emotion*, 3(4), 394-400.
- Hobson, R. (1986). The autistic child's appraisal of expressions of emotion: A further study. *Journal of Child Psychology and Psychiatry*, 27(5), 671-680.
- Hobson, R., Ouston, J., & Lee, A. (1988). Emotion recognition in autism: coordinating faces and voices. *Psychol Med*, 18(4), 911-923.
- Horn, L. (1985). Metalinguistic negation and pragmatic ambiguity. *Language*, 121-174.
- Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA. 4th edition.: Wadsworth.
- Huckvale, M. (2003). Prorec (version 1.0). *University College London*, Downloaded from <http://www.phon.ucl.ac.uk/resource/prorec/#download>.
- Huckvale, M. (2004). Speech Filing System suite (version 4.6). *University College London*, Downloaded from <http://www.phon.ucl.ac.uk/resource/sfs/>.
- Iwata, S. (1998). Some extensions of the echoic analysis of metalinguistic negation. *Lingua*, 105(1-2), 49-65.
- Izard, C. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological bulletin*, 115, 288-288.
- Jones, C., Happé, F., Baird, G., Simonoff, E., Marsden, A., Tregay, J., et al. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia*, 47(13), 2850-2858.
- Kahana-Kalman, R., & Goldman, S. (2008). Intermodal matching of emotional expressions in young children with autism. *Research in Autism Spectrum Disorders*, 2(2), 301-310.
- Kleinman, J., Marciano, P. L., & Ault, R. L. (2001). Advanced Theory of Mind in High-Functioning Adults with Autism. *Journal of Autism and Developmental Disorders*, 31(1), 29-36.
- Klin, A. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions: Biological Sciences*, 358(1430), 345-360.
- Langdon, R., Davies, M., & Coltheart, M. (2002). Understanding Minds and Understanding Communicated Meanings in Schizophrenia. *Mind & Language*, 17(1&2), 68-104.

- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *Journal of Autism and Developmental Disorders*, 30(3), 205-223.
- Loveland, K., Tunali-Kotoski, B., Chen, R., Brelsford, K., Ortegon, J., & Pearson, D. (1995). Intermodal perception of affect in persons with autism or Down syndrome. *Development and psychopathology*, 7, 409-409.
- Loveland, K., Tunali-Kotoski, B., Chen, Y., Ortegon, J., Pearson, D., Brelsford, K., et al. (1997). Emotion recognition in autism: Verbal and nonverbal information. *Development and psychopathology*, 9(03), 579-593.
- McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders*, 38(4), 325-350.
- Moll, J., de Oliveira-Souza, R., Eslinger, P., Bramati, I., Mourão-Miranda, J., Andreiuolo, P., et al. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, 22(7), 2730-2736.
- Moore, C., & Corkum, V. (1994). Social understanding at the end of the first year of life. *Developmental Review*, 14(4), 349-372.
- Noh, E. (1998). Echo questions: Metarepresentation and pragmatic enrichment. *Linguistics & Philosophy*, 21, 603-628.
- Noh, E. (2001). *Metarepresentation: A Relevance-Theoretic Approach*. Amsterdam: John Benjamins.
- Rutherford, M., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the Mind in the Voice: A Study with Normal Adults and Adults with Asperger Syndrome and High Functioning Autism. *Journal of Autism and Developmental Disorders*, 32(3), 189-194.
- Serrano, J., Iglesias, J., & Loeches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear, and surprise in 4- to 6-month-old infants. *Developmental Psychobiology*, 25(6), 411-425.
- Sperber, D. (2000). Metarepresentation in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations : a multidisciplinary perspective* (pp. 448 p.). Oxford ; New York: Oxford University Press.
- Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. In P. Cole (Ed.), *Radical Pragmatics* (pp. 295-318). New-York: Academic Press.
- Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Sutcliffe, P., & Bishop, D. (2005). Psychophysical design influences frequency discrimination performance in young children. *Journal of Experimental Child Psychology*, 91(3), 249-270.
- Takahashi, H., Matsuura, M., Koeda, M., Yahata, N., Suhara, T., Kato, M., et al. (2008). Brain activations during judgments of positive self-conscious emotion and positive basic emotion: pride and joy. *Cerebral Cortex*, 18(4), 898-903.
- Tate, A., Fischer, H., Leigh, A., & Kendrick, K. (2006). Behavioural and neurophysiological evidence for face identity and face emotion processing in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476), 2155.
- Tracy, J. L., & Robins, R. W. (2004). Putting the Self Into Self-Conscious Emotions: A Theoretical Model. *Psychological Inquiry*, 15(2), 103-125.
- Wang, A., Dapretto, M., Hariri, A., Sigman, M., & Brookheimer, S. (2001). Processing affective and linguistic prosody in autism: an fMRI study. *Neuroimage*, 13.
- Wang, A., Lee, S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain*, 129(4), 932.
- Wang, A., Lee, S., Sigman, M., & Dapretto, M. (2007). Reading Affect in the Face and Voice: Neural Correlates of Interpreting Communicative Intent in Children and Adolescents With Autism Spectrum Disorders. *Archives of General Psychiatry*, 64(6), 698.
- Weeks, S., & Hobson, R. (1987). The salience of facial expression for autistic children. *Journal of Child Psychology and Psychiatry*, 28, 137 - 151.

- Wilson, D. (2000). Metarepresentation in linguistic communication. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 411-448). Oxford: Oxford University Press.
- Wilson, D. (2006). The pragmatics of verbal irony: Echo or pretence? *Lingua*, 116, 1722-1743.
- Wilson, D. (in press). Pragmatic processes and metarepresentational abilities : The case of verbal irony. In T. Matsui (Ed.), *Pragmatics and Theory of Mind*. Amsterdam: John Benjamins.

Appendix

Note: starred items appear in all three experiments.

Stimuli included in the Manners of Speech condition

Item	Context – Target	
	Option 1 (item 1)	Option 2 (item 2)
breathless screaming	/	Ben gets to Daisy's house for dinner. He says to Daisy: <i>I bought some flowers for you!</i> Daisy wonders why Ben is breathless. (breathless) Daisy wonders why Ben is screaming. (screaming)
crying stuttering	/	Joey hears Ben in the garden. Ben says to Joey: <i>There is a dog coming!</i> Joey wonders why Ben is crying. (crying) Joey wonders why Ben is stuttering. (stuttering)
screaming whispering	/	Ben comes home. He says to his Mum: <i>I want to watch television!</i> Mum wonders why Ben is screaming. (screaming) Mum wonders why Ben is whispering. (whispering)
singing crying	/	Ally is chatting with Ben. Ben says to Ally: <i>I don't want to go to the seaside.</i> Ally wonders why Ben is singing. (singing) Ally wonders why Ben is crying. (crying)
*stuttering singing	/	Ben is in the living room. He says to his Mum: <i>I don't like watching television</i> Mum wonders why Ben is stuttering. (stuttering) Mum wonders why Ben is singing. (singing)
*whispering breathless	/	Ben meets Helen on the street. He says to Helen: <i>I want to see the football match!</i> Helen wonders why Ben is whispering. (whispering) Helen wonders why Ben is breathless. (breathless)

Stimuli included in the Physical States condition

Item	Context – Target	
	Option 1 (item 1)	Option 2 (item 2)
cold in pain	/	Ben comes back after a day skiing. He says: <i>Oh my feet!</i> Ben's feet are cold! He should have worn warmer socks. (cold) Ben is in pain. He is not used to skiing anymore (in pain)
*cold in pain	/	Ben was on a boat and fell out. He says: <i>Can you come and help me?</i> Ben is really cold. The water is freezing! (cold) Ben is in pain. He hit his head on the boat. (in pain)
*in pain tired	/	Ben's been waiting in front of the hospital for a long time. He says: <i>I've been waiting for two hours!</i> Ben is really in pain. It hurts! (in pain) Ben is really tired. He needs to go to bed! (tired)
in pain tired	/	Ben has been working all afternoon in the garden. He says: <i>I couldn't do that every day!</i> Ben is in pain. His back is aching. (in pain) Ben is very tired. He worked in the sun for too long. (tired)
tired cold	/	Ben is in the kitchen with a friend. He tells her friend: <i>I need something to drink.</i> Ben is tired. He is falling asleep. (tired) Ben is cold. He needs to warm up. (cold)
tired cold	/	Ben is back from a long walk in the mountains. He says: <i>I am happy to be back home!</i> Ben is tired. He walked for too long. (tired) Ben is cold. He walked in the snow! (cold)

Stimuli included in the Basic Emotions condition

Item	Context – Target	
	Option 1 (item 1)	Option 2 (item 2)
*angry sad	/	Ben wants to talk to Fred but Fred is not home. He says: <i>Fred hasn't come home yet?</i> Ben is angry. When Fred gets back, he will be in trouble (angry) Ben is sad. He wanted to spend the evening with Fred. (sad)
*disgusted surprised	/	Ben is in the underground station. He says: <i>It smells like food.</i> Ben is disgusted. That smell makes him feel sick. (disgusted) Ben is surprised. He didn't know there was food in that station. (surprised)
*happy disgusted	/	Ben is looking at the menu in the canteen. He says: <i>We're having tomatoes for lunch</i> Ben is happy. Tomatoes are his favorite food! (happy) Ben is disgusted. He hates tomatoes! (disgusted)
sad scared	/	Ben is at a party, he sees Frank coming towards him. He says: <i>Frank is coming.</i> Ben is sad. He knows that Frank will ask him to leave the party. (sad) Ben is scared. He doesn't want Frank to bully him. (scared)
*scared angry	/	Ben hears a big noise from his neighbours' house. He says: <i>What is that noise?</i> Ben is scared. There might be a burglar in his neighbours' house! (scared) Ben is angry. He doesn't like it when his neighbours are too noisy. (angry)
*surprised happy	/	Ben is looking at his diary. He says: <i>I'm meeting with Tom today</i> Ben is surprised. He had forgotten that he was meeting Tom. (surprised) Ben is happy. He is really looking forward to meeting Tom. (happy)

Stimuli included in the Social Emotions condition

Item	Context – Target	
	Option 1 (item 1)	Option 2 (item 2)
*guilty proud	/	Ben is in the kitchen with his Mum. He says: <i>I finished all the pasta!</i> Ben feels guilty. He shouldn't steal food. (guilty) Ben is proud of himself. He never manages to finish his food. (proud)
guilty proud	/	Ben's Mum asks him about his day at school. Ben says: <i>I fought with Mark.</i> Ben feels guilty. He shouldn't fight. (guilty) Ben feels proud. He defended the one who was bullied. (proud)
*proud sorry	/	Ben has just received his exam results. He says: <i>I came second.</i> Ben is so proud. He worked so hard for that exam. (proud) Ben is sorry. He wanted to be 1st. (sorry)
proud sorry	/	Jane's Mum wonders why Jane is not with Ben. Ben says: <i>I told her to walk home from school!</i> Ben is proud. He thinks he had a great idea! (proud) Ben is sorry. He forgot he was supposed to pick her up! (sorry)
sorry / guilty	/	Ben's dog jumps from the table onto the sofa. Ben says: <i>I taught him that trick.</i> Ben is sorry. It's quite annoying. (sorry) Ben feels guilty. Now the dog has broken a precious vase. (guilty)
*sorry guilty	/	Clare wonders who gave her phone number to Charles. Ben says: <i>I gave your number to Charles.</i> Ben is sorry. He didn't know Clare wanted her number to remain secret. (sorry) Ben feels guilty. He should have listened to Clare. (guilty)

Stimuli included in the Second Order Mental States condition

Item	Context – Target	
	Option 1 (item 1)	Option 2 (item 2)
*admiration irony	/	Glenn tells Phil that he decided to come by plane rather than by train. Ben says: <i>How clever of you!</i> Ben really thinks that Glen was quite right because the trains are always late. (admiration) Ben actually thinks that Glenn is silly because the plane takes longer than the train. (irony)
*echoic quest. endorsing att.	/	Clara says that the film was fantastic. Ben says: <i>Fantastic</i> Ben disagrees with Clara, he didn't think the film was fantastic. (echoic question) Ben also thought the film was fantastic. (endorsing att.)
echoic quest. endorsing att.	/	Vincent says that the meal was lovely. Ben says: <i>Lovely</i> Ben also thought the meal was lovely. (endorsing att.) Ben disagrees with Vincent, he didn't think the meal was lovely. (echoic question)
*admiration irony	/	Dan tells Ben that he is cooking pasta tonight. Ben says: <i>What a brilliant idea!</i> Ben is actually not very excited because he is bored of pasta! (irony) Ben is really excited because Dan cooks pasta so well! (admiration)
metaling. neg. negation	/	Ben got a book for christmas. He says: <i>I'm not happy!</i> In fact, Ben is thrilled! (metalinguistic negation) In fact, Ben is really sad! (negation)
metaling. neg. negation	/	Ben has just dropped Lisa off at the station. He says: <i>I'm not sad.</i> In fact, Ben is very sad! (metalinguistic negation) In fact, Ben is really happy! (negation)
*sincere opposite	/	Tristan has gone to the cinema without Ben. Ben says: <i>I'm not angry at all.</i> Ben is really not angry, he had other plans anyway. (sincere) Actually, Ben is angry but he doesn't want to say it. (opposite)
sincere opposite	/	Steve goes to visit his friend Ben who was sick yesterday. Ben says: <i>I'm OK don't worry.</i> Actually, Ben is in pain but he doesn't want to say it. (opposite) Ben is really feeling better now. (sincere)

		TD participants		Participants with ASD		
		Mean (SD)	Range	Mean (SD)	Range	t(df); p
Experiment 1 N = 34	Age	14;2 (1;7)	11;7-16;9	13;8 (1;11)	11;1-17;10	t(32) = -.86 ; p = .40
	BPVS score	100 (13)	83-128	106 (20)	78-145	t(32) = 1.06 ; p = .30
Experiment 2 N = 40	Age	13;10 (1;2)	12;7-16;2	13;8 (1;4)	11;5-16;3	t(38) = -.44 ; p = .67
	BPVS score	108 (11)	86-125	110 (23)	72-145	t(38) = .38 ; p = .71
Experiment 3 N = 32	Age	13;10 (1;1)	12;2-16;2	13;11 (1;4)	11;5-16;3	t(30) = .34 ; p = .74
	BPVS score	112 (18)	86-146	118 (23)	78-159	t(30) = .85 ; p = .40

Table 1. Participants' age and BPVS score.

		TD	AS	t (dl) ; p
Exp. 1 N = 34	Intensity – Mean (SD)	2.43 dB (1.62 dB)	2.43 dB (1.62 dB)	t(32) = .38 ; p = .70
	Duration – Mean (SD)	40 ms (24 ms)	72 ms (64 ms)	t(32) = 1.76 ; p = .09
	Frequency – Mean (SD)	112 Hz (82 Hz)	122 Hz (92 Hz)	t(32) = .99 ; p = .33
Exp. 2 N = 40	Intensity – Mean (SD)	2.43 dB (1.89dB)	2.43 dB (1.35 dB)	t(38) = .32 ; p = .75
	Duration – Mean (SD)	48 ms (32ms)	32 ms (24 ms)	t(38) = -1.84 ; p = .07
	Frequency – Mean (SD)	92 Hz (62 Hz)	72 Hz (72 Hz)	t(38) = -.62 ; p = .50
Exp. 3 N = 32	Intensity – Mean (SD)	2.70 dB (1.62 dB)	2.43 dB (1.08 dB)	t(29) = -.53 ; p = .59
	Duration – Mean (SD)	40 ms (32 ms)	40 ms (48 ms)	t(29) = .07 ; p = .95
	Frequency – Mean (SD)	92 Hz (62 Hz)	92 Hz (102 Hz)	t(29) = -.03 ; p = .97

Table 2. Mean threshold values (and standard deviations) for the Intensity, Duration and Frequency tasks.

Figure Captions

Figure 1. Proportion of correct answers (left) and reaction times (right) as a function of type of prosodic cue and group.

Figure 2. Proportion of correct answers (left) and reaction times (right) as a function of the type of prosodic cue and the group.

Figure 3. Number of detected Ts as a function of the Actual number of Ts; results of Post hoc Tukey tests comparing performance across conditions (n.s. – $p = n.s.$; * – $p < .05$; ** – $p < .005$).

Figure 4. Proportion of correct answers (left) and reaction times (right) as a function of the type of prosodic cue and the group.

Figures

Figure 1

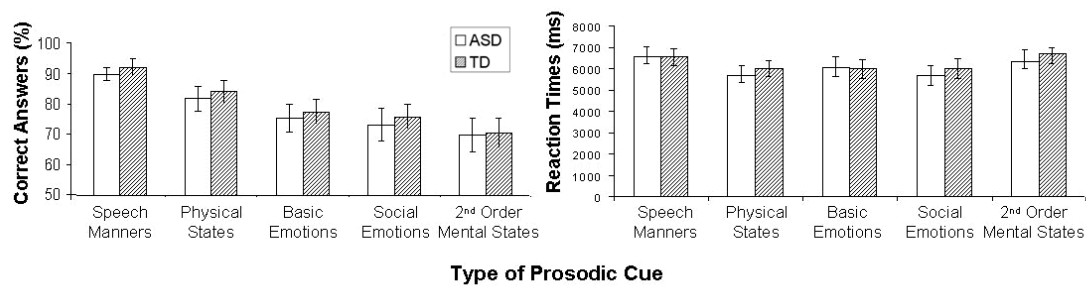


Figure 2

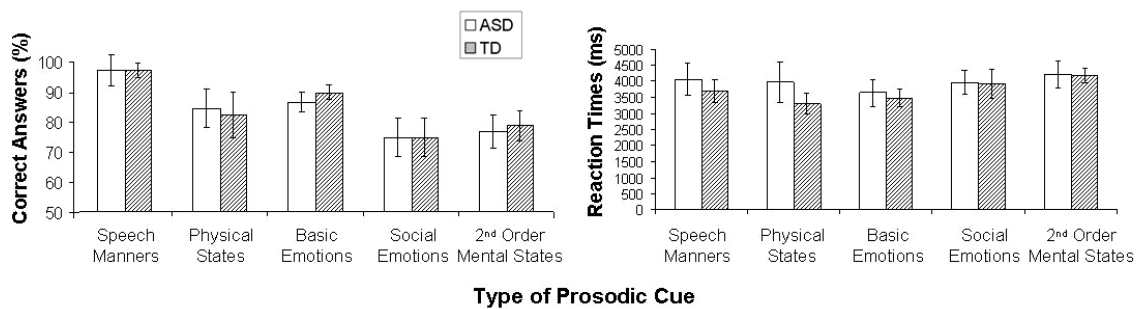


Figure 3

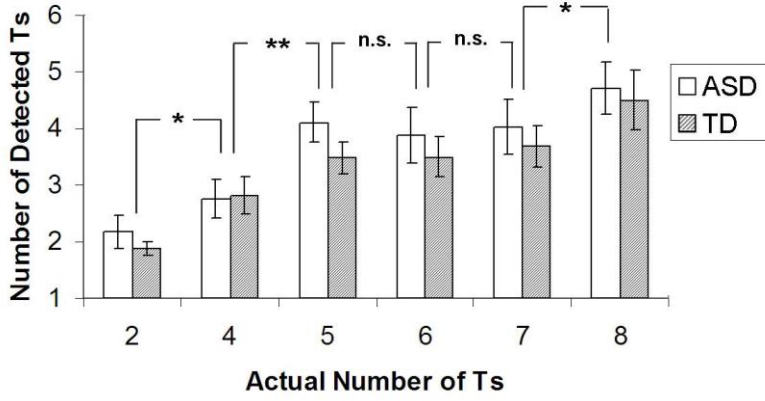


Figure 4

