Title page

Title:

A distinction between linguistic- and social-pragmatics helps the precise characterisation of pragmatic challenges in children with Autism Spectrum Disorders and Developmental Language Disorder.

Key-Words: Autistic Spectrum Disorders (ASD), Developmental Language Disorder (DLD), Theory of Mind (ToM), Structural language, Pragmatics, Grammar

Contact details (affiliation and address):

Clara Andrés-Roqueta

Affiliation: Department of Developmental, Educational, Social and Methodological Psychology. Universitat Jaume I de Castellón.

Address: Av/ Sos Baynat, s/n, 12071. Castellón, Spain.

E-mail: candres@uji.es

Telephone: + 34 964 729559

Fax: + 34 964 729262

Napoleon Katsos

Affiliation: Department of Theoretical and Applied Linguistics, English Faculty Building. University of Cambridge.

Address: TR-24, CB3 9DP, Cambridge, UK.

E-mail: nk248@cam.ac.uk

Telephone: + 44 (0) 1223 767396

Fax: + 44 (0) 1223 767398

Declaration of interests

The authors declare that they have no conflict of interest.

Funding

Funding: This study was funded by Conselleria de Educación, Cultura y Deporte of Generalitat Valenciana (Spain) (GV/2015/092), Jaume I University of Castellón (UJI-A2016-12) and a British Academy Small Project (SG-47135).

Research involving human participants and/or animals

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

Informed consent

Informed consent was obtained from all individual participants included in the study. Approvals from the local government and the ethics committee of the University Jaume I de Castellón were obtained. Schools and parents/guardians also provided informed consent.

Abstract

Purpose: Children with Autistic Spectrum Disorders (ASD) and children with Developmental Language Disorder (DLD) face challenges with pragmatics, but the nature and sources of these difficulties are not fully understood yet. The purpose of this study was to compare the competence of children with ASD and children with DLD in two pragmatics tasks that place different demands on ToM and structural language.

Method: Twenty Spanish-speaking children with ASD, twenty with DLD and forty age-and language-matched children with neurotypical development were assessed using two pragmatics tasks: a linguistic-pragmatics task, that requires competence with structural language; and a social-pragmatics task, that requires competence with Theory of Mind (ToM) as well.

Results: For linguistic-pragmatics, the ASD group performed similarly to the DLD and language-matched groups, and performance was predicted by structural language. For social-pragmatics, the ASD group performed lower than the DLD and the language-matched groups, and performance was predicted both by structural language and ToM.

Conclusions: Children with ASD and children with DLD face difficulties in linguistic-pragmatic tasks, in keeping with their structural language. Children with ASD face exceptional difficulties with social-pragmatic tasks, due to their difficulties with ToM. The distinction between linguistic- and social-pragmatic competences can inform assessment and intervention for pragmatic difficulties in different populations.

Introduction

Pragmatic ability is typically defined as the skill to use language appropriately for social and situational contexts (Geurts & Embrechts, 2008). Consequently, it relies upon the linguistic ability to produce and understand well-formed utterances, but also it relies upon the cognitive and social skills that are needed to be aware of (and respond to) the different qualities of the context (e.g. as regards the interlocutor, his/her social status, mental states or interests) (Eigsti, de Marchena, Schuh, & Kelley, 2011).

Difficulties with pragmatics are consistently reported in children with Autism Spectrum Disorders (ASD), but they have been also reported in children with other neurodevelopmental disorders, including children with Developmental Language Disorders (DLD). However, the extent of impairment in each population and the underlying causes are under debate (Andrés-Roqueta & Katsos, 2017; Norbury, 2014).

As regards ASD, having social communication difficulties (including pragmatic problems) is a diagnostic criterion for ASD (DSM-5, American Psychiatric Association, 2013). Qualitative studies report atypical and inappropriate conversational behaviour (Volden & Phillips, 2010, p. 204), and screening instruments consistently report deficits in a wide range of pragmatic skills (Geurts & Embrechts, 2008; Philofsky, Fidler, & Hepburn, 2007; Volden & Phillips, 2010). Similarly, research based on experimental tasks has described difficulties: in non-literal language understanding, like figurative language (e.g., idioms and metaphors, Happé, 1993, 1994; Norbury, 2004, 2005) or indirect requests (Ozonoff & Miller, 1996); in noticing violations of conversational maxims (Surian, Baron-Cohen, & Van der Lely, 1996); in disambiguating words with different meanings using context information (Brock, Norbury, Einav, & Nation, 2008; Jolliffe & Baron-Cohen, 1999); in making inferences from narratives and humour comprehension (Ozonoff & Miller, 1996); and in conversational skills like topic maintenance (Volden & Phillips, 2010). However, although people with ASD have been

classically described as having a uniform pragmatic impairment, recent empirical studies suggest that some areas of pragmatic competence are preserved, at least within a subpopulation of individuals. For example, Chevallier, Wilson, Happé and Noveck (2010) showed that adolescents with ASD were as able as peers with neurotypical development to draw scalar inferences associated with logical terms such as "or" (e.g. they interpreted "*There is a house or a lion*" with the inference that there is a house or a lion, *but not both*). Similar findings are reported in Pijnacker, Hagoort, Buitelaar, Teunisse, and Geurts (2009) paper, who studied a related but distinct type of scalar inference. Moreover, a study by Deliens, Papastamou, Ruytenbeek, Geelhand and Kissine (2018) also showed that adolescents and young adults with ASD were as able to peers with neurotypical development to comprehend conventionalised indirect requests through interrogative utterances (e.g. *Can you _?* constructions). Therefore, the empirical evidence suggests that some pragmatic skills can be preserved in some individuals with ASD at levels that are comparable to those of peers with neurotypical development.

Classically, pragmatic language difficulties in ASD have been attributed to other deficits that are at the base of the disorder, including incorrect or atypical mental states attribution (Theory of Mind, ToM: Baron-Cohen, Leslie, & Frith, 1985), executive dysfunctions (Hill, 2004), a weak ability to integrate information from context (Central Coherence, Happé & Frith, 2006), or lack of social motivation (Dawson et al., 2004). Each of these explanations targets a crucial aspect of pragmatic competence: without ToM a listener does not have access to the beliefs of the speaker that may help the listener reach the speaker's intended interpretation (e.g. when the speaker ironically says 'How polite you are' in a situation where the speaker patently believes the opposite); without executive control, the listener may lack the processing resources needed for much of pragmatic reasoning (e.g. she may have difficulties inhibiting a literal response and activating alternative interpretations); weak

ability to use information from the context will lead to interpretations that miss out on critical aspects of the situation or speaker cues; and without social motivation, people may be deprived of social inputs and learning opportunities. These may, on their own may, lead to lower ability to understand other people's intentions, but may also do so indirectly, by leading to diminished expertise in social cognition (Dawson et al., 2004).

However, as pointed by Geurts and Embrechts (2008), pragmatic ability might also depend on structural language skills (such as vocabulary and syntax), that can fluctuate from non-verbal to above average levels in school-age children with ASD (Volden, Coolican, Garon, White, & Bryson, 2009). Children with ASD who develop age-appropriate structural language also show difficulties with pragmatics (Volden et al., 2009). This alternative proposal has not yet been fully investigated, but there is increasing evidence that some of the pragmatic language deficits observed in children with ASD are due to difficulties with their structural language. For example, Norbury (2005) reported that children with ASD had difficulties with metaphor comprehension but these difficulties were predicted by their grammar and vocabulary skills. Moreover, structural language (but not ToM) was a significant predictor of success with figurative language understanding. In the same vein, a recent meta-analytic review of 41 experimental studies in figurative language (including metaphor) showed that the statistical differences observed between groups of children with neurotypical development and ASD disappeared when the groups were matched on the language level (Kalandadze, Norbury, Nærland, & Næss, 2016). Similarly, better structural language skills predicted better sensitivity to violation of pragmatic maxims in participants with ASD (Chevallier et al., 2010; Pijnacker et al., 2009).

Turning to DLD (the new term to replace Specific Language Impairment, SLI), it is an impairment characterised by a pronounced delay in the acquisition of structural language (vocabulary and/or grammar) but with relative strengths in pragmatics, in the absence of

other primary deficits (e.g. in social motivation). However, studies using screening instruments (Norbury, Nash, Baird, & Bishop, 2004) and conversational analysis (see Adams, 2002, for an overview) do report pragmatic deficits among DLD population. Similarly, research from experimental tasks describes problems in: figurative language understanding (Norbury, 2005); sensibility to conversational maxims like "quantity" (Davies, Andrés-Roqueta, & Norbury, 2016; Katsos, Andrés-Roqueta, Estevan, & Cummins, 2011; Surian et al., 1996); using preceding context to resolve ambiguous utterances (Brock et al., 2008); and in narrative production, especially when it comes to providing relevant and essential information to understand a story, and to descriptions of internal states like emotions, beliefs or intentions of the characters (Norbury, Gemmell, & Paul, 2014). Norbury et al. (2014) observed that children with DLD had more difficulty not only than their peers with neurotypical development, but also than peers with ASD, in both relevance and internal states. The authors propose that children with DLD may be able to infer the mental states of other people, but they may not have the specific vocabulary to talk about thoughts, emotions or intentions.

Most of the empirical evidence shows that the difficulties with pragmatics in children with DLD tend to be in keeping with the children's levels of structural language and not with their ToM skills (e.g. false belief understanding, see for example Davies et al., 2016; Gernsbacher & Pripas-Kapit, 2012; Norbury, 2005 or Volden et al., 2009), and they tend to perform as well as younger children with neurotypical development matched on language level on pragmatic tasks (Katsos et al., 2011). However, the development of general structural language and ToM are closely related from early childhood on, and children with DLD are sometimes at risk for delays in ToM development around critical periods of acquisition, especially when pragmatic deficits are present (Miller, 2006). This provides additional

evidence supporting the idea that pragmatic deficits may be in keeping with the level of structural language.

To take stock, the discussion above documents that there are some discrepancies in the literature on whether children with ASD do face difficulties with in some areas of pragmatics or not. Moreover, the nature of those deficits is also in question: there are different views on whether any such difficulties are due to the lower structural language skills that children with ASD often have (in which case the nature of pragmatic difficulties might not be different to the nature of difficulties with pragmatics that children with DLD face) or whether the difficulties are due to an intrinsic aspect of ASD, such as challenges with ToM, Central Coherence or social motivation.

As argued by Andrés-Roqueta and Katsos (2017), it is possible that variation in which independent variables (such as ToM and structural language) have been assessed in each study can account for the differences in findings with regards to core causes of pragmatic difficulties in people with ASD. Furthermore, the authors point that an additional factor that may explain the different conclusions reached in previous findings is that the different tasks used to measure pragmatics skills may involve structural language and ToM to a greater or lesser extent. A differentiation between kinds of pragmatic inferences has been supported by theoretical pragmatics literature taking the ToM-related demands of the tasks as the critical variable to categorize them (e.g. Deliens et al., 2018; Kissine, 2016; O'Neill, 2012), suggesting that sometimes an utterance can be fully understood based on pragmatic norms and the context from ones' egocentric perspective, without the need to infer others' perspective (see Andrés-Roqueta & Katsos, 2017, for a detailed explanation). For example, Deliens et al. (2018) compared the performance of adults with ASD and neurotypical adults in two pragmatic skills: interpretation of indirect requests and the comprehension of irony. They found that adults with ASD could understand indirect requests, but they showed

difficulties in understanding irony. The authors concluded that preserved pragmatics in people with ASD was limited to those cases where egocentric processing of context (that does not rely on ToM) was sufficient for correct interpretation.

While the specific details of irony processing are a topic of much investigation (see Pexman, 2008), the relevant observation is that without attributing to the speaker a belief that is incompatible with what was literally said, the participant will not be able to understand and subsequently explain the meaning of the speaker's utterance. Empirical evidence shows important associations between ToM competence and irony understanding. Specifically, correlational studies in children with neurotypical development and children with ASD have shown that irony comprehension is related to False Belief understanding (Filippova & Astington, 2008; Happé, 1993, respectively), and also correlations have been observed between irony and higher scores in the social skills subscale of the Autism Quotient in neurotypical adults (Spotorno & Noveck, 2014). Moreover, recent studies in neuro-typical adults show that ToM reasoning and irony understanding activate similar neural regions (Spotorno, Koun, Prado, Van Der Henst, & Noveck , 2012).

In the present study, we will use the terms *linguistic-pragmatics* and *social-pragmatics* to classify the pragmatic tasks used, as suggested by Andrés-Roqueta and Katsos (2017). On the one hand and as pointed by the authors, *linguistic-pragmatics* tasks mainly require structural language and pragmatic competence to perform successfully. Take for example, sensitivity to informativeness in a quantifier comprehension task (e.g. understanding that when five out of five boxes shown to the participant have a certain object inside them, it is better say that "all" boxes have an object, than that "some" boxes have an object inside them). Here, participants need to use their pragmatic competence to avoid being under-informative (respecting the Gricean maxim of quantity, Grice (1975): "some" is an under-informative term in this case) together with their structural language (grammatical and vocabulary skills). Moreover, when

the situation is visually depicted (i.e. the participant sees five out of five boxes containing an object inside them), the information that is needed in order to evaluate if a utterance is under-informative is visually accessible and shared by the experimenter who utters the sentence and the participant who understands it, and therefore ToM skills might be less necessary than in other types of pragmatic tasks. Therefore, scalar inference tasks (such as the task used by Katsos et al., 2011 in DLD population; or the task used by Chevallier, et al., 2010, and by Pijnacker, et al., 2009 in ASD populations), would be considered *linguistic-pragmatics* tasks. The task used to assess the understanding of indirect requests by Deliens et al. (2018) would also fall under the category of *linguistic-pragmatics*, because the meaning of expressions such as 'can you…' is standardised by convention and frequency of use and its understanding does not require on-the-spot reasoning about speaker's intentions.

On the other hand, *social-pragmatics* tasks include those cases where in addition to structural language and pragmatic competence, it is necessary to use ToM abilities (e.g. identifying hidden intentions under non-literal language, such as in the Strange Stories task originally created by Happé, 1994). This task includes stories where a character uses ironic and sarcastic utterances like "*Well, that's very nice, isn't it! That is what I call politeness!*" to remark that a person is being *rude*. So, in order to understand the intended meaning (*you are being rude*), the participant must use his/her structural language competence to understand the literal meaning, as well as an understanding of the pragmatic maxims that instruct interlocutors to be truthful and relevant. But in contrast with the *linguistic-pragmatics* task, the participant needs to make a correct inference of the beliefs of the speaker (*the speaker believes that the other person is not being polite*). By inferring correctly the speaker's belief, the listener can inhibit attributing to the speaker a belief which –even though it appears to be coherent with what is said *("…that's very nice… That is what I call politeness!*")- is nevertheless factually unreasonable and false. Instead, by correctly inferring the speaker's

belief, the listener infers the factually reasonable belief ("*that is rude*"). This belief appears to be incoherent with the literal meaning of the utterance, but becomes coherent once the correct pragmatic inference about irony is made. For similar reasons, we would classify the irony task used in Deliens et al. (2018), as well as tasks that require the understanding of humour or sarcasm, particular in their social aspects (Silva, Da Fonseca, Esteves, & Deruelle, 2017), as a case of *social-pragmatics*. The implication of the distinction between linguistic- and social-pragmatics is that pragmatic competence will be a function of the linguistic- and ToM profile of the children but also of the demands that the specific pragmatic task places on structural language and ToM. Given that in ASD populations there can be a dissociation of structural language skills and ToM, some of the previously puzzling findings about pragmatics in ASD could be explained by this distinction. Indicatively, the participants with ASD tested by Chevalier et al. (2010) and Pijnacker et al. (2009) on informativeness in conditions where there was no need to engage ToM, had language skills within the typical range and did not face challenges with pragmatics compared to their aged-matched peers with neurotypical development. Similarly in Deliens et al. (2018) study, participants with ASD preformed similar as their aged-matched peers with neurotypical development on a task about conventional use of language (which is associated to vocabulary and grammar), but they performed worse on the irony task (where ToM was needed). More generally, we would expect that pragmatic tasks that require the attribution of intentions, such as irony, sarcasm, and humour –which we would classify as *social-pragmatics*- will depend on the ToM skills of people with ASD and people with DLD; while tasks where there is no need to engage ToM –which we would classify as *linguistic-pragmatics*- will depend primarily on their level of structural language.

In the present study we directly test the hypothesis that the pragmatic language difficulties of children with ASD and children with DLD will depend on their structural language skills in

linguistic-pragmatic tasks, but will additionally depend on their ToM skills in social-pragmatic tasks.

Specifically, as a first hypothesis it is expected that language-matched children with ASD and DLD will face significant difficulties in the *linguistic-pragmatics* task, and to the same extent. Moreover, both clinical groups are expected to face more difficulties than an age-matched group with neurotypical development. However, their performance will not fall below that of a language-matched group with neurotypical development.

Furthermore, structural language (but not ToM) will be a significant predictor of successful performance for both clinical groups in the *linguistic-pragmatic* task.

As a second hypothesis it is expected that both groups will also face significant difficulties in the *social-pragmatics* task in comparison to the age-matched group with neurotypical development. However, children with ASD will face exceptional difficulties relative to the children with DLD and the language-matched group with neurotypical development, due to anticipated deficits in ToM skills.

Moreover, in addition to structural language, ToM will be a significant predictor of success for both clinical groups in the *social-pragmatics* task.

We must note that between-group differences in structural language and ToM are not a main research question of the present paper, and we are only specifically interested in between-group differences on the *linguistic-pragmatic* and the *social-pragmatic* task.

Methods

Participants

All the children who took part in the study, including those with ASD or DLD, were native Spanish speakers and attended public schools in Spain. At the time of the study, all participants with ASD or DLD were receiving language and communication intervention by a

speech and language therapist at school while attending mainstream classrooms, following the Spanish legislation for inclusive education.

According to the aims of the study, children were classified in four groups: children with ASD (N = 20), children with DLD language-matched to each ASD child on receptive grammar (N = 20), participants with neurotypical development matched for chronological age (AM, N = 20) and younger participants with neurotypical development matched for language to each child with ASD on receptive grammar (LM, N = 20).

Recruitment procedure and group characteristics

Educational psychologists and speech and language therapists from the schools helped the research group with recruiting 4- to 10-year-old children meeting a current diagnosis of ASD or DLD according to regional county diagnostic protocols. From the potential clinical sample, the researchers recruited those who were native speakers of Spanish, whose speech was intelligible, and who had no hearing loss, intellectual impairment or highly gifted background.

Children with ASD and DLD had a confirmed and updated diagnosis by a qualified psychologist prior to our study. All the psychologists who made the diagnoses of the children in our sample were members of the same local health services and they used the same regional county diagnostic protocols, ensuring that there was uniformity in the procedures and criteria used for the diagnoses.

ASD*:* The final ASD group comprised 20 children (5 girls, 15 boys).

- *Diagnosis of ASD*: children were recruited if they had a confirmed diagnosis of ASD based on DSM-5 criteria at the time of the study (American Psychiatric Association, APA, 2013). Their records confirmed that they all met full criteria for autism based on: the Spanish Inventory for Autistic Spectrum *IDEA* (*Inventario de espectro autista*, Rivière, 2002) and/or the Spanish version of the Social Communication Questionnaire

(SCQ: Rutter, Bailey, & Lord, 2005), for first detection; and the Autism Diagnostic Interview-Revised using both the diagnostic and current behaviour algorithms (ADI-R: Rutter, Le Couteur, & Lord, 2006), and the Autism Diagnostic Observation Schedule (ADOS, Lord, Rutter, DiLavore, & Risi, 2001), for a current diagnosis of ASD. All these measures were part of children's medical and educational records.

DLD*:* The DLD group included 20 children (6 girls, 14 boys).

- *Diagnosis of DLD:* children were recruited if they had a history of language delay in their clinical record; they did not have sensorineural hearing loss, or any other known mental or physical disorder; also, no medical condition likely to affect language was reported in their clinical record, such as a diagnosis of ASD. Their records confirmed that they had significant language disability as the primary reason for receiving speech and language therapy in the presence of typical non-verbal intelligence, based on standardized tests: a) General linguistic measures, such as the *Batería del Lenguaje Objetivo y Criterial-Screening Revisado* (BLOC-SR, Puyuelo, Renom, Solanas, & Wiig, 2007) or the Peabody Picture Vocabulary Test (PPVT-III: Dunn, Dunn, & Arribas, 2006); and b) General cognitive and neuropsychological tests, such as the Wechsler Intelligence Scale for Children (WISC-IV: Wechsler, 2005) or the Spanish "Cuestionario de Madurez Neuropsicológica Escolar" (CUMANES: Portellano, Mateos, & Martínez, 2010).
- *Confirmation of DLD by research group:* inclusion in the DLD group was conditional on scoring 1SD below age-appropriate level on either of two standardized Spanish language measures administered by the research group: a receptive grammar test, the "Comprensión de Estructuras Gramaticales" test (CEG; Mendoza, Carballo, Muñoz, & Fresneda, 2005), which is the Spanish version of the Test for Reception of Grammar for English-2 (TROG-2; Bishop, 1989); and an expressive grammar test, the "Sentence

Recall" subtest of the 'Evaluación del Lenguaje Infantil' battery (ELI; Saborit & Julián, 2005), which measures expressive language ability and short term auditory memory.

In addition, to further confirm the absence of substantial ASD traits in the DLD group, children were screened by the research group for autistic behaviour in accordance with the guidelines suggested in the Spanish version the Children's Communication Checklist-2 (CCC-2; Bishop, 2003; Spain/Spanish - Version 2 of 22 Jun 12 - MAPI Institute). This test is designed to measure language abilities, but it also includes two subscales tapping current autistic behaviour: social interaction and specific interests (subscales *I* and *J*). Previous studies with similar samples (e.g. Norbury, 2005) have established the following normative means for this measure: 22-23 for children with neurotypical development, around 11-12 for children with DLD, and below 5 for children with ASD. In our sample, the group of children with DLD scored a mean of 15.21 (SD = 5.31, range 5–24), which is between the scores for children with neurotypical development and children with DLD in previous studies. No participant with DLD scored below 5, which would have suggested presence of ASD (see Norbury et al., 2004, for detailed information of how to interpret the means). However, a single DLD participant scored exactly 5. However, this participant was 4;1 years-old (the lowest age-range at which the CCC-2 can be administered is 4;0). At such a young age, some of the features categorised as ASD by the CCC-2 may still be appropriate for children with neurotypical development (e.g. interest in playing with some particular objects). Furthermore, CCC-2 scores were unavailable for 5 children in the DLD group, as the parents of the children did not attend an interview. Overall, the high means and the range of scores from the CCC-2 subscales on current autistic behaviour of the DLD group corroborate the absence of an ASD diagnosis that was evident in the clinical and educational records of the children.

In line with previous research (Katsos et al., 2011), CEG raw scores were used to match each participant with DLD to each participant with ASD, as this test was originally developed to investigate particular aspects of DLD. Gender-matching with the children with ASD was achieved except for one case.

Chronological age matched (AM): The AM group included 20 children (6 girls, 14 boys). They were matched to children with ASD within ± 3 months of age. Gender matching was achieved, except for one case.

Language-matched (LM): The LM group included 20 children (5 girls, 15 boys). Each participant with ASD was also matched on the raw score of the CEG with a participant with neurotypical development of the same gender.

Teachers and school-based educational psychologists ensured that the children recruited for the groups of children with neurotypical development did not have any communication impairments or other neurodevelopmental disorders.

Participants in all four groups were recruited only if they scored above the 5th percentile on a non-verbal reasoning measure administered by the research group (Coloured Progressive Matrices, CPM; Raven, Raven, & Court, 1998).

Assessments

Non-verbal measures.

All the participants were assessed with a non-verbal reasoning measure (Coloured Progressive Matrices, CPM; Raven et al., 1998), with the aim of ensuring that their non-verbal IQ was within the typical range.

Language measures.

The following standardized language measures for Spanish were administered to create the *Language composite score*:

- *Grammar (receptive):* The CEG test evaluates the ability to select the matching picture for sentences of increasing grammatical complexity (Mendoza et al., 2005). Score range: 0 - 80.

- *Grammar (expressive):* The subtest *Sentence Recall* from the standardized language battery *Evaluación del Lenguaje Infantil* (ELI; Saborit & Julian, 2005) assesses expressive grammar and memory. Score range: 0 - 10.

- *Vocabulary (receptive):* The receptive vocabulary subscale of the ELI assessed the ability to identify objects and concepts. Score range: 0 - 30.

- *Vocabulary (expressive):* The expressive vocabulary subscale of the ELI assessed the ability to name objects, people and places. Score range: 0 - 30.

Language composite score: the four language measures were used to examine language as a complex construct. However, as these measures use different scoring ranges, a composite score was created. Raw scores were used, rather than standardized scores, as they are a direct indicator of how many correct responses each child achieved in each test. First, we confirmed that the four measures were highly correlated within the sample of 80 children ($r = .60$, $p < .001$). Then, the measures were weighted equally and combined to form the *language composite score*. This approach was taken to address issues arising from different scaling within language measures and between the other non-standardized measures used in the study. Specifically, the *language composite* variable was obtained by adding together all four of the linguistic raw scores. The final score ranged from 0 to 100 with each language measure representing ¼ of the new composite score.

A note about the reliability and validity of the Spanish language measures used:

The CEG and ELI have been used in previous research in order to assess language competence in Spanish children with DLD (e.g. Katsos et al., 2011). Both tests meet the standards for reliability and validity: the CEG test has appropriate psychometric properties (Reliability: $\alpha = .91$, see Mendoza et al. 2005 for further details) and good construct validity

with other existing measures (correlations ranging from .12 - .80, see Muñoz-López, Fresneda, Mendoza-Lara, & Carballo, 2008 for further details). The ELI scales have also shown adequate psychometric properties (Sentence Recall: α = .71; Expressive vocabulary: α = .83; Receptive vocabulary: α = .70; Pragmatics: α = .84) and construct validity of all subscales (correlation with other existing measures) ranging from .51 - .85 (see Saborit & Julian, 2005 for further psychometric details).

Theory of mind measures.

Two first-order false-belief measures were used to assess the ability to understand another person's mental states, and therefore to create the ToM composite score:

- *Unexpected Content* task (see Perner, Leekam, & Wimmer, 1987, for detailed information). Children were shown a tube of Smarties and they were asked what was inside the tube. After the children answered, the tube was opened to show it actually contained a small pencil. The experimenter then closed the tube of Smarties and asked a control question: "*What does the tube really contain?*". After that, the experimenter asked the first test question: ''*When you first saw the tube, what did you think was inside it?*'' (*own false belief*). Finally, the experimenter asked about another person's (who was outside the room) false belief: *"What would … think is inside the tube?"* (*3rd person false belief question*). Recalling one's own false belief is also an important aspect of ToM (see Wellman, Cross, & Watson, 2001 for a review). So, it requires not a report of a different person's mental states at current time, but of one own's mental states, albeit at a different point in time.

Responses were scored with 0.5 points when the 3rd person's false-belief was correctly predicted and 0.5 points when one's own false belief was correctly predicted giving the same weight to both answers, following the scoring of previous studies (see Farrar et al., 2006 for further details). Score range: 0 – 1. The answers to both false-belief questions were coded as correct only if the participant had also correctly answered the control question.

- *Change of Location* task (see Wimmer & Perner, 1983, for detailed information).

Children were told a story (while looking at pictures) about Sally and Ann, in which Sally leaves a ball in a basket. Then Sally goes outside the room and Ann puts the ball in a different location (a box). Then, Sally comes back to play with her ball. At this moment the experimenter asked the test question: "*Where will Sally look for her ball?*" (Sally's false belief). Control questions were asked to check if the children remembered where Sally had put the ball initially ("*Where did Sally put her ball before leaving the room?*") and its location after Ann puts the ball to a box ("*Where is it now?*"). As all the participants were Spanish-speaking children tested in Spain, we used two Spanish names for the main characters (instead of Sally and Ann).

Responses were scored with 1 point when the 3rd person's false-belief was correctly predicted. The answers to the false-belief questions were coded as correct only if the participant had also correctly answered both control questions.

ToM composite score: The scores from the two tasks were added to produce a composite score (score range: 0 - 2).

Pragmatic measures

- *Linguistic-pragmatics: Cavegirl task.*

This task measures sensitivity to the pragmatic maxim of informativeness. Children are placed in front of a computer and the experimenter explains that they are going to play a game with a Cavegirl (a fictional character) who wants to learn better Spanish. During the game they are shown visual displays on the screen, while listening to the Cavegirl saying how many toys are in the boxes. Specifically, she describes the contents of the boxes using expressions of quantity such as '*all'*, '*none'*, '*some'* (or *some not* or *not all*) and '*most*'. Descriptions with '*all'* included visual screens where 5 out of 5 or 2 out of 5 boxes had an object inside. Descriptions with '*none'* included visual screens where no boxes or 2 out of 5

boxes had an object inside. Descriptions with '*some'*, '*some… not'* and *'not all'* included screens where 0, 2, or 5 out of 5 boxes had an object inside. Finally, descriptions with '*most'* included screens where 2, 4, or 5 out of 5 boxes had an object inside. If what the Cavegirl says is right, children are instructed to tell her ''*that is right*''; whereas if what she says is wrong, the child should tell her ''*that is wrong*'' (see Katsos et al., 2011, for detailed task information).

From these combinations, the following conditions can be created:

1) When the statement was logically true (e.g. statements with 'all', 'none', 'some', 'not all' made with visual screens depicting 5/5, 0/5, 2/5 and 2/5 boxes respectively having an object)

2) When the statement was logically false (e.g. statements with 'all', 'none', 'some', 'not all' made with visual screens depicting 2/5, 2/5, 0/5 and 5/5 boxes respectively having an object;

3) When the statement was logically true but pragmatically under-informative (e.g. statements with 'some' and 'not all' made with visual screens depicting 5/5 and 0/5 boxes respectively having an object).

In the critical condition for informativeness (condition 3), participants hear a description that is semantically true but pragmatically under-informative (e.g. in a situation where *all* the books are inside the boxes, they hear the statement '*some* of the books are inside the boxes'). This statement is literally true, because at least a subset of the books have the property of being inside the boxes). Participants who are sensitive to the maxim of informativeness should reject these statements on the grounds that the statement is under-informative (in this example, the Cavegirl should have said 'all'). In addition to Katsos et al. (2011), who used the under-informativeness paradigm with children with DLD, Chevalier et al. (2010) and Pijnacker et al. (2009) used it to study the pragmatic competence of people with ASD and

more than twelve studies to date have used it with children with neurotypical development and adults. In the present study, we only used the 24 trials from the Cavegirl task (those regarding condition 3) where the correct response required pragmatic competence (in this paper, we call it the *Ling-Pragmatic* variable for later analyses): 6 in a situation where '*all'* would have been the informative description (because the visual screen shows 5/5 objects having an object inside them) but the Cavegirl used '*some'*, 6 where '*all'* would have been informative but she used '*most'*, 6 where '*none'* would have been informative but she used *'not all'* and another 6 where '*none'* would have been informative but she used '*some…not'*. The necessary skills to succeed in this task are: receptive structural language skills, to understand the meaning of each sentence, and pragmatic competence, to judge if an utterance is informative or not in the specific situation where it is used. A participant may employ ToM to represent the Cavegirl's knowledge of the situation, namely how many objects she believes are inside the boxes. However, representing the Cavegirl's knowledge is not necessary in order to determine whether or not her utterance was informative. The knowledge that is critical for success in the task (how many objects are inside the boxes in each specific situation) is obtained through visual information that is available to the participant via direct perception.

A note about the reliability and validity of the Cavegirl task:

Condition 3 of the task used in the present study was seen to have appropriate reliability within our sample (24 trials: 6 questions x 4 quantifiers, $N = 80$, $\alpha = .944$). Moreover, the Cavegirl task has been used in previous research in order to investigate competence with the expressions for "all", "none", "some", "some…not", and "most" in 31 languages by testing 768 children and 536 adults (see Katsos et al., 2016 for further details). The authors found a cross-linguistically similar order of acquisition of quantifiers across different languages by

using this task, which gives it construct validity through its use in different samples and languages.

- *Social pragmatics: Strange Stories.* Six stories from the *Strange Stories* task originally designed by Happé (1994) that are age-appropriate for the participants in our sample were used: Pretence, Joke, Lie, White Lie, Irony and Figure of Speech (see O'Hare, Bremner, Nash, Happé, & Pettigrew, 2009, for further details about each story). As all the participants were Spanish-speaking children, we used Spanish names for the main characters of the selected stories (instead of the names of the characters in the original task).

This task is typically used to assess mentalizing through the recognition of the communicative intentions of people using indirect or non-literal utterances. We call this task the *Social-Pragmatic* variable for later analyses, because as the name of the task itself suggests, the protagonists have unusual or unexpected intentions (based on beliefs and desires), which are not compatible with the literal meaning of what they say. Accessing these mental states correctly is a prerequisite for making the pragmatic inferences that render what the protagonists say appropriate for the context. So, inferring these beliefs and desires requires mentalizing in each specific context. We emphasize that while we are using the Strange Stories task because of the mentalizing component that is a prerequisite for the pragmatically appropriate understanding of what was said, the participants' responses are coded for pragmatic appropriateness (relevance of the answer given to the question that was asked), accuracy (informativeness), and veracity (truthfulness of information provided), rather than for mentalizing. This scoring system is in line with the scoring used in previous research (e.g. Freed et al., 2015; O'Hare et al., 2009), but in the scoring proposed in the present study different pragmatic aspects (relevance, informativeness and truthfulness) can be dissociated, e.g., an answer may justify the protagonist's utterance based on mental states, but

the mental states mentioned may be irrelevant to explain the actual intention of the main character.

Questions and answers were in Spanish originally. The stories were presented in random order. Only the answers to justification questions related to what was said (*Why does he/she say this/that?*) were scored. The final score was obtained by adding appropriateness, accurateness, and veracity scores (range: 0 - 3): 3 points (explicit intention or psychological state: appropriate, accurate and truthful), 2 points (implicit intention or psychological state: appropriate and truthful), 1 point (partially explicit intention: accurate) or 0 points (incorrect: nor appropriate, accurate or truthful). Appropriateness, accuracy and veracity were scored along the following lines:

- Appropriateness (relevance): Does the answer show the child has understood the implicit intention of the speaker appropriately? That is: the answer is coherent with the question, and it provides relevant/pertinent information to answer it.
- Accuracy (informativeness): Does the answer give as much information as needed to explain satisfactorily the implicit intention? That is: the answer provides enough information about the mental states of the characters.
- Veracity (quality): Does the answer give information that is factually true given the evidence? That is: the answer includes only information supported by the evidence in the story and does not include other information not related to the story or not supported by the evidence.

Two indicative examples of the adapted scoring are presented in Appendix A.

In total, the first author and a research assistant blind to the children's diagnosis and the hypotheses of the study scored 480 responses (6 stories x 80 participants) independently. Disagreements between coders were resolved through discussion. A new score was mutually agreed in cases of disagreement. Inter-scorer-reliability was high (Cohen's kappa: .96).

The necessary skills to succeed are: receptive structural language in order to understand each story; expressive structural language, to explain the protagonist's response; pragmatic competence, to judge if what the main character said is appropriate in each context; and ToM, to represent the characters' mental states in order to give a coherent explanation. ToM skills are extensively required and play a greater role in Strange Stories than in the Cavegirl task as outlined above.

Procedure

Approvals from the local government and the ethics committee of the University Jaume I de Castellón were obtained. Schools and parents/guardians also provided informed consent. Children were tested individually during school-time. In the first two sessions (30 minutes each) we administered non-verbal reasoning, linguistic, and ToM measures. The pragmatic tasks (60 minutes) were administered in one subsequent session. The order of presentation within the session was counterbalanced to avoid order effects.

Data analyses

Data analysis was conducted using the statistical package SPSS (version 24). When the sample was subdivided into four groups (ASD, DLD, AM and LM), the data failed the Shapiro–Wilk test of normality, showing several unequal variances across groups for language, ToM and pragmatics scores. This suggests the need to be appropriately cautious regarding the results presented in the following section because it was convenient to use less rigorous methods of statistical analysis, given the lack of normality of the data. Kruskal–Wallis chi-squared tests were used to investigate main effects, and Mann–Whitney U (two-tailed, significance threshold of .05) to examine differences between groups on key measures. Effect sizes of group comparisons were calculated using *r* using the formula: $r = (z) / (\sqrt{N})$, because according to Fritz, Morris, and Richler (2012), when between-group comparisons are made with Mann-Whitney U tests, size effects must be calculated using "*r*" and not "*d*".

According with Fritz et al. (2012), a value of 0 - .1 is considered a small effect; .2 - .4 is considered a medium effect; and .5 - 1 is considered a large effect.

Moreover, zero-order nonparametric correlations (Spearman) between key measures were conducted separately in the ASD and DLD groups. Finally, in order to further investigate the contribution of age, language skills, ToM and type of developmental disorder (ASD/DLD) on the pragmatics tasks, a hierarchical linear regression analysis was conducted for the whole clinical sample (N = 40). When investigating the status of the groups as regards the presence of a developmental disorder (clinical vs neurotypical development) the data failed the Shapiro–Wilk test of normality and a bootstrapping method was implemented using 1000 bootstrap samples for deriving robust estimates of standard errors, confidence intervals, and p-values of the regression model. As pointed by Zhu and Jing (2019), in the absence of a normal distribution, the Bootstrap method can improve the stability of regression coefficients and reduce the length of confidence intervals.

Results

Descriptive statistics and between-group comparisons on key and related measures

Table 1 reports the descriptive statistics of the four groups for grammar, age, language, ToM, and the two pragmatic measures.

Table 1 here

Matching variables (receptive grammar and age) and variables of interest (language and ToM)

The Kruskal-Wallis test showed that the groups differed in age H (3, N = 80) = 19.99, p < .001. Post-hoc pairwise comparisons revealed significant differences in age between LM-ASD (U = 60.50, p < .001, r = .597) and LM-DLD (U = 85.000, p = .002, r = .492), and large and medium effect sizes were observed; but no difference in age between ASD-DLD (U

= 182.000, p = .626, r = .077), ASD-AM (U = 191.500, p = .818, r = .036) or DLD-AM (U = 168.500, p = .394, r = .135).

As regards the matching variable (raw scores in CEG test), the Kruskal-Wallis test showed significant differences between groups H (3, N = 80) = 25.73, p < .001. Post-hoc pairwise comparisons revealed that ASD and DLD groups did not differ on the receptive grammar matching variable (U = 188.000, p= .745, r = .052). Moreover, neither of the clinical groups differed from the LM group on this (ASD-LM: U = 169.000, p = .401, r = .133; DLD-LM: U = 175.500, p = .507, r = .232). In contrast, the AM group out-performed LM (U = 67.500, p < .001, r = .567), ASD (U = 46.000, p < .001, r = .659) and DLD groups (U = 37.000, p < .001, r = .698), and large size effects were observed in all cases.A highly significant effect of group was found on the language composite, H (3, N = 80) = 24.215, p < .001). The ASD and DLD groups did not differ significantly between them (U = 174.000, p = .482, r = .085) or from the LM group (ASD-LM: U = 188.500, p = .756, r = .017; DLD-LM: U = 175.000, p = .499, r = .104). In contrast, the AM group outperformed the ASD and DLD groups and large effect sizes were found (ASD-AM: U = 50.500, p < .001, r = .670; DLD-AM: U = 60.000, p < .001, r = .614).

A main effect of group was observed in ToM scores (H (3, N = 80) = 16.404, p = .001). Children with ASD had lower ToM scores when compared to the AM group (U = 65.500, p < .001, r = .605), the DLD (U = 106.000, p = .008, r = .417) and the LM groups (U = 113.000, p = .014, r = .387). Large and medium size effects were observed. In contrast, children in the DLD group did not differ significantly from the AM (U = 152.000, p = .140, r = .233) or the LM groups (U = 197.500, p = .942, r = .011), and small size effects were found.

Pragmatic measures

Turning to the linguistic-pragmatic task (the Cavegirl task on informativeness), a main effect of group was found, H (3, N = 80) = 18.120, p < .001). Between-group comparisons revealed

no reliable differences between the ASD and the DLD group (U = 176.000, p = .515, r = .103) or between the ASD and the LM group (U = 180.500, p = .596, r = .084). Similarly, no significant differences were found between the DLD and LM group (U = 168.000, p = .384, r = .138). The AM group outperformed both the ASD (U = 77.000, p = .001, r = .532) and the DLD group (U = 76.000, p = .001, r = .538), and large effect sizes were observed in both cases.

Turning to the social-pragmatics task (Strange Stories), again a main effect of group was found, H (3, N = 80) = 17.35, p = .001. Between-group comparisons showed that the ASD group performed significantly lower than DLD group (U = 100.000, p = .006, r = .431), and a medium effect size was observed. The ASD group also performed lower than LM and AM groups: ASD-LM (U = 103.000, p = .008, r = .418) and a medium size effect was observed, and ASD-AM (U = 72.000, p < .001, r = .552) with a large size effect found. The DLD group scored lower than the AM group (U = 124.000, p = .039, r = .326), but at the same levels as the LM group (U = 198.000, p = .957, r = .008). In this case, a medium and a small effect size were observed respectively.

Note about DLD group:

Since five out of the twenty children in the DLD group did not have CCC-2 scores on current autistic behaviour, we conducted a comparison on ToM tasks and the two pragmatic tasks between the group of children with DLD who had CCC-2 scores (DLD-1, N=5) and the group of children with DLD children who did not (DLD-2, N=15). The two subgroups with DLD were not significantly different to each other in ToM scores (DLD-1: N = 5, mean = 1.40, SD = .894, range: 0 – 2; DLD-2: N = 15, mean = 1.267, SD = .776, range: 0 – 2), and a small size effect was found (U = 33.500, p = .735, r = .084). Regarding the two pragmatic tasks, similar between-group comparisons revealed no significant differences between them neither in the *linguistic-pragmatics* task (U = 23.500, p = .230, r = .274), nor in the *social-*

pragmatics task (U = 26.500, p = .349, r = .217), and medium size effects were observed. Descriptive statistics for *linguistic-pragmatics* task were the following (DLD-1: N = 5, mean = .60, SD = .389, range: .4 – 1; DLD-2: N = 15, mean = .342, SD = .314, range: .4 – 1), and for *social-pragmatics* task the following (DLD-1: N = 5, mean = 9.20, SD = 6.30, range: 2 – 16; DLD-2: N = 15, mean = 6.27, SD = 4.132, range: 2-15).

Correlations between pragmatic measures, age, language and ToM.

Zero-order nonparametric correlations (Spearman) between key measures are presented in Table 2. For children with ASD, a medium and positive correlation was observed between Linguistic-pragmatics and Language (p = .037). Moreover, Social-pragmatics was positively and strongly correlated with Language (p = .007) and ToM (p = .001).

For children with DLD, a medium and positive correlation was observed between Linguistic-pragmatics and Language (p = .012); and Social-pragmatics was positively and strongly correlated with all the variables: Age (p < .001), Language (p < .001) and ToM (p = .002). Moreover, medium and positive correlations between the two pragmatic measures were observed within both the ASD (p = .025) and DLD groups (p = .026).

Table 2 here

Predictive analysis of the pragmatic measures

As presented in the correlation analyses, many of the variables are intercorrelated in ASD and DLD groups, making it difficult to identify the independent contribution each one makes to the pragmatic measures. For example, language is significantly correlated with ToM which begs the question: are the False Belief tasks that we used just a sophisticated language test (a possibility raised by Norbury, 2004), or do they (additionally) measure social cognition? To further investigate the contribution of these variables, two hierarchical linear regression analyses using a bootstrap method were conducted for the whole clinical sample (ASD + DLD, N = 40).

Linguistic- and social-pragmatics were the outcome variables in the regression, and four predictor variables were entered in the following order (Table 3): age was entered first, because raw scores had been used and there was an important age difference between some participants in the sample (age ranging from 4- to 10-years-old), with older participants expected to score higher. Language was entered next, because structural language deficits are thought to be fundamental factors for pragmatic deficits observed in children (Norbury, 2004). ToM scores were entered after language to investigate if ToM has a specific contribution to each type of pragmatic competence when structural language skills have been taken into account. The variable DLD v ASD diagnosis status was introduced in the final stage as a binary covariate (Reference group = ASD), to check if other features related to an ASD or DLD diagnosis are relevant.

In order to explore whether the relationship between ToM and Language to each pragmatic measure was different or similar in the clinical groups, the interactions between DLD v ASD status and Language, and DLD v ASD status and ToM were further explored. However, these interactions were not significant and they were not further explored in the linear regression analyses.

For Linguistic-pragmatics, the general model was significant and accounted for a total of 27% of the variance, $F (4, 35) = 3.222$, $R^2 = .269$, $p = .024$. Higher scores on Linguistic-pragmatics were positively and significantly associated with higher Language scores, which explained 17% of the variance. No unique association was found with Age, ToM or the diagnostic variable (ASD v DLD).

For Social-pragmatics, the general model was also significant, and accounted for a total of 61% of the variance: $F (3, 35) = 13.599$, $R^2 = .608$, $p < .001$. After Age was entered in the first step without being a significant predictor, results showed that Language and ToM were

both significant predictors of successful performance (explaining 16% and 20% of the variance, respectively). Again, the diagnostic variable (ASD v DLD) was not significant.

Table 3 here

Discussion

In the present study we capitalised on a distinction between linguistic- and social-pragmatics made in Andrés-Roqueta and Katsos (2017) to test if pragmatic competence will be a function of the linguistic and ToM profile of children with ASD and DLD, but also of the demands that the specific pragmatic task places on structural language and ToM. In this regard, we aimed to test the hypothesis that the pragmatic difficulties of children with ASD and children with DLD would be in keeping with their structural language skills in tasks that require *linguistic-pragmatic* competence, but will additionally depend on their ToM skills in tasks that require *social-pragmatic* competence.

It was first hypothesized that language-matched children with ASD and DLD would face significant and similar difficulties in the *linguistic-pragmatics* task (in this study, the Cavegirl task, a task that measures sensitivity to the pragmatic maxim of informativeness, by Katsos et al., 2011), and both clinical groups were expected to face more difficulties than an age-matched group with neurotypical development. However, their performance would not fall below that of the language-matched group of children with neurotypical development. Also, structural language (but not ToM) was expected to be a significant predictor of successful performance for both clinical groups. Between-group comparisons confirmed that children with ASD and children with DLD faced significant difficulties on the *linguistic-pragmatics* task compared to age-matched peers with neurotypical development. Difficulties with *linguistic-pragmatics* were in keeping with the children's structural language and the two clinical groups scored as well as the language-matched peers with neurotypical development. Moreover, regression analyses revealed that structural language but not ToM was a significant predictor of success

in the two groups with developmental disorders, highlighting the role of structural language in the *linguistic-pragmatics* task. In this sense, the first hypothesis was verified, and structural language skills predicted the performance of clinical groups on our linguistic-pragmatic task (in this study, sensitivity to informativeness), reaching similar findings of previous research (e.g. Chevallier et al., 2010; Davies et al., 2016; Kalandadze et al., 2016; Norbury, 2005; Pijnacker et al., 2009).

Secondly, regarding the *social-pragmatics* task (in this study, the Strange Stories test by Happé, 1994, a task that measures recognition of the communicative intentions under indirect or non-literal utterances, but in the present study we modified the scoring of the answers according to appropriateness to context), it was hypothesized that both clinical groups would also face significant difficulties in comparison to the age-matched group with neurotypical development. However, due to anticipated deficits in ToM skills children with ASD were expected to face exceptional difficulties in comparison with children with DLD or younger children with neurotypical development with similar language level. Moreover, in addition to structural language, ToM was expected to be a significant predictor of success for both clinical groups. In this regard, between-group comparisons showed that children with ASD and children with DLD faced difficulties with the *social-pragmatics* task but to a different extent. Children with DLD, whose ToM and structural language skills were comparable to those of the LM group, performed as well as the LM group. Children with ASD, whose ToM scores were lower than the LM and DLD groups, performed significantly lower than these groups, even though all three groups were matched on structural language level. Highlighting the role of ToM in *social-pragmatics* tasks, regression analyses revealed that both structural language and ToM were significant predictors of success in the two clinical groups. In this sense, our second hypothesis was also verified, and ToM was also seen a good predictor of the *social-pragmatics* task, according to previous studies with similar tasks (Filippova &

Astington, 2008; Happé, 1993; Spotorno et al., 2012). However, it should be noted as a limitation that this statement is true for the combined clinical group (ASD + DLD). Even with the addition of diagnosis as a factor in the predictive analysis, the fact that this relationship may not hold for the ASD group on its own or for the DLD group on its own cannot be ruled out, particularly given that ToM and *social-pragmatics* were both statistically significantly lower in the ASD group than in the DLD group. Therefore, we cannot know if this relationship could be driven by group differences rather than by relationships between ToM and *social-pragmatics* within each group per separate.

Together, these findings identify structural language and ToM as key skills for pragmatics in the clinical groups. In no case were other features of ASD or DLD (which were captured under the binary diagnostic label variable, DLD v ASD) a significant predictor of success. Nevertheless, and related to with the limitation stated in the previous paragraph, it is important to point out that the fact no other features of ASD or DLD were being significant predictors of success for any of the pragmatic tasks does not mean that the relationships observed hold for each clinical group separately.

The different contribution of ToM and structural language in pragmatics tasks observed in the present study is compatible with some of the most intriguing findings coming out from recent studies on children with ASD. For example, Deliens et al. (2018) reported that adults with ASD and verbal IQ skills within the typical range faced challenges with understanding some aspects of pragmatics (irony) but not others (indirect speech acts). The authors' preferred explanation, namely that the former requires taking the perspective of the speaker while the latter does not, captures the essence of the distinction that the current study supports.

Moreover, turning to some conflicting findings, such as Norbury (2005) and Happé (1993) who reached different conclusions as regards the contribution of ToM skills in metaphor, our research suggests that it is important to comprehensively assess structural language skills, an

idea that is highlighted as the main conclusion of the meta-analysis by Kalandadze et al. (2016). However, by showing that there are also aspects of pragmatics where structural language is not the sole predictor (our *social-pragmatics* task based on Strange Stories), the results of the present study suggest that structural language is a crucial but not the only significant predictor of pragmatic competence (e.g. see Happé's, 1993 irony experiment), above all in those situations where pragmatic inferences require taking into account an implicit communicative intention of the speaker. It must be noted though that measuring ToM only with false belief tasks captures just one single aspect of ToM (first order false beliefs), rather than ToM comprehensively. In future studies it would be interesting to explore this relationship by addressing more ToM areas (e.g. affective and cognitive ToM), to see which specific aspects of mental state understanding are related to *linguistic* and *social-pragmatics*.

In sum, a conclusion that could be derived from the present study is that the extent and underlying causes of pragmatic difficulties of children with ASD and DLD depend not only on the children's competence with structural language and ToM, but also on the specific task used to measure pragmatics, as pointed by Andrés-Roqueta and Katsos (2017).

With the presentation of different kinds of pragmatics that require different types of skills, this research challenges two widely-held views about pragmatics in ASD pointed by Andrés-Roqueta and Katsos (2017). First, that children with ASD are exceptionally challenged by pragmatics. In the present study, children with ASD do show difficulty with pragmatics across the board in comparison with age-matched peers with neurotypical development, but –in the case of *linguistic-pragmatics*- these deficits are commensurate with their structural language skills rather than posing exceptional challenges. And second, that these challenges are only due to deficits in ToM or social motivation which are intrinsic to ASD. Instead, we show that when an increasingly important distinction between types of pragmatics is taken into account, children with ASD can do as well as younger children with neurotypical development with the

same language level in *linguistic-pragmatics* tasks, with structural language serving as the key predictor of success.

The present paper addresses a relevant area of research, as most research in the domain of pragmatics has focused on the unitary construct of pragmatics. Our findings do emphasize the importance of correctly conceptualizing the competences that are necessary for success with a pragmatics task, and in particular the need to understand the extent of impairment in children with ASD and DLD, and the underlying causes (in the present study, the role of ToM and structural language), a point made clearly by Adams (2002) and Norbury (2014) as well as more recently (Andrés-Roqueta & Katsos, 2017). As demonstrated in the present study, structural language skills and ToM affect pragmatic ability. Therefore, focusing exclusively on pragmatics without taking into account other language and cognitive skills will not reveal the actual communication needs of a child with neurodevelopmental disorders and might result in too narrowly defined treatment goals (Geurts & Embrechts, 2008). Consequently, multi-disciplinary assessments of the social communication profile of a child are necessary to design an adapted intervention for each child (e.g. include structural language assessments together with pragmatic ones, to find out if the pragmatic difficulties are in keeping with structural language level or not). Moreover, including structural language contents in interventions that target pragmatic competence (both in social and linguistic scenarios) may be profitable for almost all children who show pragmatic difficulties as pointed by Kalandadze et al. (2016). And finally, it would be interesting to find out if there are differences in performance between types of pragmatic (*linguistic* and *social*) in real conversational contexts, which is of major importance for intervention (Chevallier et al., 2010).

As a limitation, even though the children of the sample were carefully matched, the age-range was wide and other important changes apart from ToM and structural language development also take place between 4 - 10 years of age (e.g. frontal lobe development and development

of executive function, which arguably will also play a role in succeeding at pragmatic tasks; Hill, 2004). Future studies should address this issue to complete the picture of pragmatic difficulties across children with ASD and other neurodevelopmental populations, using a wide range of pragmatic tasks as well.

Moreover, we must be cautious with our results given the lack of normality of the sample that took part in this study. A greater sample size is needed in future studies to carefully study the overlap between metalinguistic, structural language and different pragmatic skills as they link to ToM, and also to study how these relationships hold for each diagnostic group on its own. It must be noted that an oral response (and subsequent expressive language skills) was required for the *social-pragmatic* task (Strange Stories task), but it was not required for the *linguistic-pragmatic* task (Cavegirl task). This issue could contribute to differences observed between both tasks, and future studies must address this issue by designing pragmatic experiments with similar expressive language demands, above all when participants have communication disorders. Moreover, the precise characterisation of the distinction between *linguistic*- and *social-pragmatics* and whether it depends on the phenomenon under study (e.g. informativeness or irony), on the interpretative strategies of the listener (Kissine, 2016), or the situational context (Andrés-Roqueta & Katsos, 2017), requires further investigation.

Acknowledgements

This study was funded by Conselleria de Educación, Cultura y Deporte of Generalitat Valenciana (Spain) (GV/2015/092), Jaume I University of Castellón (UJI-A2016-12) and a British Academy Small Project (SG-47135). ***Declaration of interest:*** The authors report no conflicts of interest. They are solely responsible for the content and form of the paper.

References

Adams, C. (2002). Practitioner review: the assessment of language pragmatics. *Journal of Child Psycholy and Psychiatry, 43*(8), 973-87. https://doi.org/10.1111/1469-7610.00226

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Andrés-Roqueta, C., & Katsos, N. (2017). The contribution of grammar, vocabulary and theory of mind in pragmatic language competence in children with autistic spectrum disorders. *Frontiers in Psychology,* 8, 996. https://doi.org/10.3389/fpsyg.2017.00996

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition, 21*(1), 37-46. https://doi.org/10.1016/0010-0277(85)90022-8

Bishop, D. V. (1989). *Test for the reception of grammar: (TROG)*. Medical Research Council.

Bishop, D. V. (2003). *The Children's Communication Checklist: CCC-2*. London: Harcourt Assessment.

Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, *108*(3), 896-904. https://doi.org/10.1016/j.cognition.2008.06.007

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar inferences in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *40*(9), 1104-1117. https://doi.org/10.1007/s10803-010-0960-8

Davies, C., Andrés-Roqueta, C., & Norbury, C. F. (2016). Referring expressions and structural language abilities in children with specific language impairment: A pragmatic tolerance account. *Journal of Experimental Child Psychology*, *144*, 98-113. https://doi.org/10.1016/j.jecp.2015.11.011

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., & Estes, A. (2004). Defining the early social attention impairments in autism: Social orienting, joint attention, and responses to

emotions. *Developmental Psychology, 40,* 271–283. https://doi.org/10.1037/0012-1649.40.2.271

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective pragmatic impairment in autism spectrum disorder: Indirect requests versus irony. *Journal of Autism and Developmental Disorders, 48*(9), 2938-2952. https://doi.org/10.1007/s10803-018-3561-6

Dunn, L. M., Dunn, L. M., & Arribas, D. (2006). *PPVT-III Peabody: test de vocabulario en imágenes*. TEA Ediciones.

Eigsti, I. M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders, 5*(2), 681-691. https://doi.org/10.1016/j.rasd.2010.09.001

Farrar, M. J., Johnson, B., Tompkins, V., Easters, M., Zilisi-Medus, A., & Benigno, J. P. (2009). Language and theory of mind in preschool children with specific language impairment. *Journal of Communication Disorders, 42*(6), 428-441. https://doi.org/10.1016/j.jcomdis.2009.07.001

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79,* 126-138. https://doi.org/10.1111/j.1467-8624.2007.01115.x

Freed, J., McBean, K., Adams, C., Lockton, E., Nash, M., & Law, J. (2015). Performance of children with social communication disorder on the Happé Strange Stories: Physical and mental state responses and relationship to language ability. *Journal of Communication Disorders, 55,* 1-14. https://doi.org/10.1016/j.jcomdis.2015.03.002

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2. https://doi.org/10.1037/a0024338

Gernsbacher, M. A., & Pripas-Kapit, S. R. (2012). Who's missing the point? A commentary on claims that autistic persons have a specific deficit in figurative language comprehension. *Metaphor and Symbol, 27*(1), 93-105. https://doi.org/doi.org/10.1080/10926488.2012.656255

Geurts, H. M., & Embrechts, M. (2008). Language profiles in ASD, SLI, and ADHD. *Journal of autism and developmental disorders, 38*(10), 1931-1943. https://doi.org/10.1007/s10803-008-0587-1

Grice, H. P. (1975). *Logic and conversation.* In P. Cole, & J. Morgan (Eds.), Syntax and semantics (pp. 41-58). New York: Academic Press.

Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48,* 101-119. https://doi.org/10.1016/0010-0277(93)90026-R

Happé, F. G. (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. https://doi.org/10.1007/BF02172093

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5-25. https://doi.org/10.1007/s10803-005-0039-0

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences, 8*(1), 26-32. https://doi.org/doi.org/10.1016/j.tics.2003.11.003

Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: linguistic processing in high-functioning adults with autism or Asperger syndrome: is local coherence impaired?. *Cognition, 71*(2), 149-185. https://doi.org/10.1016/S0010-0277(99)00022-0

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K. A. B. (2016). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism, 22*(2) 99-117. https://doi.org/10.1177/1362361316668652

Katsos, N., Andrés-Roqueta, C., Estevan, R. A. C., & Cummins, C. (2011). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification?. *Cognition*, *119*(1), 43-57. https://doi.org/10.1016/j.cognition.2010.12.004

Katsos, N., Cummins, C., Ezeizabarrena, M-J., Gavarró, A., Kraljević, J. K., Hrzica, G., ... Özge, D. (2016). Cross-linguistic patterns in the acquisition of quantifiers. *Proceedings of the National Academy of Science of the United States of America*, *113*(33), 9244-9249. https://doi.org/10.1073/pnas.1601341113

Kissine, M. (2016). Pragmatics as Metacognitive Control. *Frontiers in Psychology*, *6*, 1-11. https://doi.org/10.3389/fpsyg.2015.02057

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2001). *Autism Diagnostic Observation Schedule*. Los Angeles, CA: Western Psychological Services.

Mendoza, E., Carballo, G., Muñoz, J., & Fresneda, M. D. (2005). *CEG, Test de comprensión de estructuras gramaticales*. Madrid: TEA Ediciones.

Miller, C. A. (2006). Developmental relationships between language and theory of mind. *American Journal of Speech-Language Pathology, 15*, 142–154. https://doi.org/10.1044/1058-0360(2006/014)

Muñoz-López, J., Fresneda, M. D., Mendoza-Lara, E., & Carballo, G. (2008). Propiedades psicométricas de una prueba de comprensión gramatical. *Revista de Neurología, 47*(1), 21-23. https://doi.org/10.33588/rn.4701.2008037

Norbury, C. F. (2004). Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1179-1193. https://doi.org/10.1044/1092-4388(2004/087)

Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology*, *23*(3), 383-399. https://doi.org/10.1348/026151005X26732

Norbury, C. F. (2014). Practitioner review: social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *Journal of Child Psychology and Psychiatry, 55*(3), 204-216. https://doi.org/10.1111/jcpp.12154

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: a validation of the Children's Communication Checklist—2. *International Journal of Language & Communication Disorders*, *39*(3), 345-364. https://doi.org/10.1080/13682820410001654883

Norbury, C., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: a cross-disorder comparison. *Journal of Child Language*, *41*(03), 485-510. https://doi.org/10.1017/S030500091300007X

O'Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal of Autism and Developmental Disorders*, *39*(6), 916-928. https://doi.org/10.1007/s10803-010-0971-5

O'Neill, D. K. (2012). Components of pragmatic ability and children's pragmatic language development. *Cognitive pragmatics, 4*, 261-287.

Ozonoff, S., & Miller, J. N. (1996). An exploration of right-hemisphere contributions to the pragmatic impairments of autism. *Brain and Language*, *52*(3), 411-434. https://doi.org/10.1006/brln.1996.0022

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year old's difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*, 125-137. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Pexman, P. M. (2008). It's fascinating research: the cognition of verbal irony. *Curent Directions in Psychological Science*, *17*, 286-290. https://doi.org/10.1111/j.1467-8721.2008.00591.x

Philofsky, A., Fidler, D. J., & Hepburn, S. (2007). Pragmatic language profiles of school-age children with autism spectrum disorders and Williams syndrome. *American Journal of Speech-Language Pathology, 16*(4), 368-380. https://doi.org/10.1044/1058-0360(2007/040)

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J. P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 39*(4), 607-618. https://doi.org/10.1007/s10803-008-0661-8

Portellano, J. A., Mateos, R., & Martínez, R. (2010). *Cuestionario de Madurez Neuropsicológica Escolar (CUMANES)*. Madrid: TEA Ediciones.

Puyuelo, M., Renom, J., Solanas, A., & Wiig, E. H. (2007). *BLOC-SR: Batería del Lenguaje Objetivo y Criterial-Screening Revisado*. Madrid: TEA Ediciones.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Coloured Progressive Matrices*. Oxford: Oxford Psychologists Press.

Rivière, Á. (2002). *IDEA: inventario de espectro autista*. Buenos Aires: Fundación para el desarrollo de los estudios cognitivos (FUNDEC).

Rutter, M., Bailey, A., & Lord, C. (2005). *SCQ: Cuestionario de comunicación social: manual*. Madrid: TEA Ediciones.

Rutter, M., Le Couteur, A., & Lord, C. (2006). *Entrevista para el Diagnóstico del Autismo revisada (ADI-R)*. Madrid: TEA Ediciones.

Saborit, C., & Julián, J. P. (2005). *L'avaluació del llenguatge infantil. ELI.* Universitat Jaume I de Castellón. Colección: Educación.

Silva, C., Da Fonseca, D., Esteves, F., & Deruelle, C. (2017). Seeing the funny side of things: Humour processing in Autism Spectrum Disorders. *Research in Autism Spectrum Disorders, 43*, 8-17. https://doi.org/https://doi.org/10.1016/j.rasd.2017.09.001G

Spotorno, N., & Noveck, I. A. (2014). When is irony effortful? *Journal of Experimental Psychology: General, 143*(4), 1649-65. https://doi.org/10.1037/a0036630

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage, 63*(1), 25-39. https://doi.org/10.1016/j.neuroimage.2012.06.046

Surian, L., Baron-Cohen, S., & van der Lely, H. K. J. (1996). Are children with autism deaf to Gricean maxims? *Cognitive Neuropsychiatry, 1*, 55–71. https://doi.org/10.1080/135468096396703

Volden, J., Coolican, J., Garon, N., White, J., & Bryson, S. (2009). Brief report: Pragmatic language in autism spectrum disorder: Relationships to measures of ability and disability. *Journal of Autism and Developmental Disorders, 39*(2), 388. https://doi.org/10.1007/s10803-008-0618-y

Volden, J., & Phillips, L. (2010). Measuring pragmatic language in speakers with autism spectrum disorders: Comparing the Children's Communication Checklist—2 and the Test of Pragmatic Language. *American Journal of Speech-Language Pathology, 19*(3), 204-212. https://doi.org/10.1044/1058-0360(2010/09-0011)

Wechsler, D. (2005). *Escala de Inteligencia de Wechsler para niños (WISC-IV)*. Madrid: TEA Ediciones.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta- analysis of theory- of- mind development: The truth about false belief. *Child development, 72*(3), 655-684. https://doi.org/10.1111/1467-8624.00304

Wimmer, H., &, Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 21*, 103-28. https://doi.org/10.1016/0010-0277(83)90004-5

Zhu, J., & Jing, P. (2010). The analysis of bootstrap method in linear regression effect. *Journal of Mathematics Research, 2*(4), 64-69.

Tables and Figures

Table 1. Descriptive statistics of participants and between-group comparisons on key and related measures.

	Gender (F:M)	Age	Grammar	Language	ToM	Ling-P	Social-P
ASD (A)							
M	5:15	85.70	47.20	62.09	.62	.45	3.5
SD		18.96	14.38	14.79	.76	.32	4.7
Range		48 - 120	11-66	33.74 - 88.08	0 - 2	.08 - 1	0 - 14
DLD (B)							
M	6:14	81.20	48.70	63.10	1.30	.41	7
SD		22.56	12.59	17.55	.78	.34	4.76
Range		49 - 127	17-68	24.05 - 90.37	0 - 2	.04 - 1	2 - 16
AM (C)							
M	6:14	86.20	66.15	82.88	1.65	.80	10.60
SD		19.06	8.99	9.14	.61	.28	5.32
Range		50 - 111	39-78	65.49 - 93.29	0 - 2	.08 - 1	0 - 17
LM (D)							
M	5:15	59.05	51.40	61.61	1.27	.46	7.4
SD		15.44	12.75	16.42	1.50	.23	5.07
Range		38 - 92	30-70	37.69 - 88.18	0 - 2	.04 - .96	0 - 18
H		19.99	25.73	24.215	16.404	18.12	17.35
p		< .001	< .001	< .001	.001	< .001	.001
Post hoc		*A = B* *A = C* *A > D**** *B = C* *B > D****	*A = B* *A < C**** *A = D* *B < C**** *B = D*	*A = B* *A < C**** *A = D* *B < C**** *B = D*	*A < B** *A < C**** *A < D** *B = C* *B = D*	*A = B* *A < C*** *A = D* *A < C*** *B = D*	*A < B*** *A < C**** *A < D*** *B < C** *B = D*

Note[1]: Post hoc by two-tailed Mann-Whitney U-test, where * $p < .05$; ** $p < .01$; and with *** $p < .001$.
Note [2]: *Age* = chronological age (months); *Grammar* = raw scores on CEG test (receptive grammar); *Language* = Language composite score; *ToM* = Theory of Mind composite score; *Ling-P* = Linguistic-Pragmatics; Social-P = Social-Pragmatics.

Table 2. Zero-order nonparametric Spearman correlations between critical variables within the ASD and DLD group.

	ASD (N=20)				DLD (N=20)			
	Age	**Language**	**ToM**	**Ling-P**	**Age**	**Language**	**ToM**	**Ling-P**
Language	.39	-			.86**	-		
ToM	.28	.58*	-		.58**	.42	-	
Ling-P	.08	.47*	.30	-	.42	.55*	.20	-
Social-P	.40	.58**	.67**	.50*	.77**	.82**	.66**	.50*

Note [1]: *$p < .05$; **$p < .01$

Note [2]: *Age* = chronological age (months); *Language* = Language composite score; ToM = Theory of Mind composite score; *Ling-P* = Linguistic-Pragmatics; Social-P = Social-Pragmatics.

Table 3. Summary of regression coefficients for Linguistic-pragmatic scores and Social-pragmatic scores within the DLD and ASD groups (bootstrap results based on 1000 bootstrap samples).

	Ling-P					Social-P				
Predictor	**ΔR^2**	***B***	**SE B**	***p***	**95 CI**	**ΔR^2**	***B***	**SE B**	***p***	**95 CI**
Step 1	***.085***					***.219***				
Constant		.046	.171	.783	[-.301, .388]		-4.194	2.466	.099	[-8.546, 1.217]
Age		.005	.002	.058	[.000, .009]		.113	.033	.004	[.044, .170]
Step 2	***.172***					***.160***				
Constant		-.189	.200	.338	[-.621, .189]		-7.661	2.974	.016	[-13.711, -1.964]
Age		-.001	.003	.735	[-.005, .005]		.033	.042	.432	[-.056, .109]
Language		.011	.004	.005	[.003, .018]		.162	.048	.005	[.067, .258]
Step 3	***.009***					***.199***				
Constant		-.214	.203	.287	[-.630, .175]		-5.841	2.180	.008	[-10.746, -1.623]
Age		-.001	.003	.756	[-.006, .005]		.030	.035	.373	[-.045, .094]
Language		.012	.004	.006	[.004, .020]		.091	.043	.038	[.013, .180]
ToM		-.042	.060	.496	[-.156, .076]		3.028	.811	.006	[1.425, 4.547]
Step 4	***.003***					***.030***				
Constant		-.180	.240	.447	[-.668, .289]		-7.394	2.435	.006	[-.12.287, -2.487]
Age		-.001	.003	.702	[-.007, .005]		.041	.034	.230	[-.036, .101]
Language		.012	.004	.008	[.004, .020]		.094	.041	.032	[.021, .182]
ToM		-.028	.069	.694	[-.159, .115]		2.423	1.115	.047	[.269, 4.546]
ASD v DLD		-.043	.107	.661	[-.271, .162]		1.955	1.558	.253	[-1.121, 4.902]

Note: *Ling-P* = Linguistic-Pragmatics; Social-P = Social-Pragmatics; *Age* = chronological age (months); *Language* = Language composite score; ToM = Theory of Mind composite score; ASD v DLD = binary ASD/DLD status covariate (Reference group = ASD).

Appendices

Appendix A. Examples of scoring Strange Stories answers to the question '*Why does he/she say that?'* according to pragmatic appropriateness.

Pretence Story: Maria and Rosa are playing at home. Rosa takes a banana from the fruit bowl, puts it near her ear and says to Maria: "*Look! I'm talking on the phone!*"

White lie story: It is Elena's birthday. Elena was waiting for this day to arrive because she was looking forward to having a new pair of skates. Elena wants the skates more than anything else in the world. When the time to blow out the candles comes, her friends give her a gift in a big box. Elena is sure that inside the box there is a pair of skates. But when she opens the box, she sees no skates. Instead, there is a ball! That is not the gift she is really wanted. Then, her friends ask her: "*Do you like our gift?*". And Elena replies: "*Oh yes, thanks. The ball is very nice. It is just the gift I wanted!*".

Table S1. Two examples of Scoring Strange Stories answers according to context adequation for the Social-Pragmatics variable.

Scoring	Target response	*Pretense story*	*White lie story*
3 points	Explicit intention (or psychological state) *(Appropriate, accurate, truthful)*	*Because she pretends she talks on the phone.*	*Because she wants her friends to be happy.*
2 points	Implicit intention (or psychological state) *(Appropriate, truthful)*	*Because the shape is similar.*	*Because they are her friends.*
1 point	Partially explicit intention *(Accurate)*	*Because her friend is sad and she wants to make her feel happy.*	*Because she wants to make a joke.*
0 points	**Incorrect** *(Nor appropriate, accurate, truthful)*	*I don't know; She is going to eat it; The banana is not a phone*	*Because they are not her friends; She doesn't like presents; Because she's lying* (as a set response to every story)