Research in Autism Spectrum Disorders 58 (2019) 30–38

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

journal homepage: www.elsevier.com/locate/rasd

Irony comprehension and mentalizing ability in children with and without Autism Spectrum Disorder

Ronit Saban-Bezalel[a,b], Dror Dolfin[c], Nathaniel Laor[d], Nira Mashal[b,e,*]

[a] Department of Communication Disorders, Ariel University, Israel
[b] The School of Education Bar-Ilan University, Ramat-Gan, Israel
[c] Geha Mental Health Center, Petah Tikva, Israel
[d] Department of Psychiatry, Sackler Faculty of Medicine, Tel Aviv, Israel
[e] Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel

ARTICLE INFO

Number of reviews completed is 2

Keywords:
ASD
Irony
ToM
Hinting test

ABSTRACT

Background: Despite evidence suggesting that individuals with Autism Spectrum Disorder (ASD) may have difficulties in comprehension of figurative language, and irony in particular, previous studies examining this ability among individuals with (ASD) have reported inconsistent findings, resulting in different suggested etiologies of this difficulty. In view of the inconclusive findings, the current study assesses the contribution of various factors to irony comprehension, with specific focus on the association between mentalizing ability and irony comprehension.

Method: Irony comprehension was examined in 20 individuals with ASD (age range 10–15) as compared to 20 typically developing (TD) peers (age range 10–15) through a task of reading ten ironic comics. Participants were matched by age, gender, vocabulary, executive function abilities, and their results on a second-order false-belief task. Their mentalizing abilities were examined by the Hinting Test for the ability to understand intentions.

Results: A significant difference in irony comprehension was observed between the groups. Readjusting the group matching according to the Hinting Test scores eliminated the group difference in irony comprehension. Multilevel regression with logit link analysis showed that hinting and mental flexibility contributed to irony comprehension.

Conclusions: The study's findings demonstrated that individuals with ASD showed adequate ability to comprehend irony but nonetheless were outperformed by TD peers (matched on age, language, ToM, and executive functioning abilities). A comparison of the two groups by their mentalizing abilities (through understanding intentions) revealed similar comprehension abilities. These findings highlight the importance of using several tools that each focus on different aspects of mentalizing when assessing this skill in studies of figurative language in ASD.

1. Introduction

Irony is one subtype of figurative language that is used to convey messages and criticism in an indirect and sometimes amusing way; it helps strengthen social bonds (Clark & Gerrig, 1984) and achieve communication goals (Dews & Winner, 1995; Roberts & Kreuz, 1994). Irony is common in everyday discourse (Gibbs, 2000) and is frequently used in blogs (Whalen, Pexman, Gill, & Scott, 2013). Irony comprehension requires complex mental representations in which the listener must go beyond the literal meaning of the

* Corresponding author at: School of Education Bar-Ilan University, Ramat-Gan 52900, Israel.
E-mail address: mashaln@mail.biu.ac.il (N. Mashal).

https://doi.org/10.1016/j.rasd.2018.11.006
Received 18 January 2018; Received in revised form 7 November 2018; Accepted 11 November 2018

utterance and identify the speaker's pragmatic intention within the present context. The ability to identify irony begins to develop around the age of five or six (Angeleri & Airenti, 2014), and improves over the years (Aguert, Laval, Gauducheau, Atifi, & Marcoccia, 2016). The extensive developmental course of irony acquisition suggests that it involves other skills, such as figuring out whether or not the speaker believes what he or she has said, and identifying the communicative intent and motivation of the utterance. These capabilities develop at different times, which is probably why irony takes so long acquire (Filippova & Astington, 2010; Pexman & Glenwright, 2007).

The accumulated evidence indicates difficulties in the comprehension of figurative language, and irony in particular, in individuals with ASD (Mackay & Shaw, 2004; Wang, Lee, Sigman, & Dapretto, 2006) and adults with Asperger's syndrome (Martin & Mcdonald, 2004). Consistent with these findings, adults with ASD showed poorer comprehension of ironic texts than typically developing (TD) adults matched in age, expressive vocabulary, and non-verbal IQ (Saban-Bezalel & Mashal, 2015). Difficulties in irony comprehension were also observed in children and adolescents with ASD. Mackay and Shaw (2004) investigated the comprehension of six subtypes of figurative language in those with ASD and TD individuals. The groups were matched in age and vocabulary. Of all the subtypes of figurative language, irony was the most difficult for the ASD participants. In another study, participants with ASD listened to ironic and literal scenarios with various contextual cues (prosody cues, event outcome, or both) and were asked to describe the speaker's intent (Wang et al., 2006). The participants with ASD were less accurate than TD individuals (matched for age and verbal and performance IQ).

However, some studies have shown that individuals with ASD have an adequate ability to identify and comprehend irony (Colich et al., 2012; Glenwright & Agbayewa, 2012). Furthermore, in studies that impose minimal linguistic and pragmatic demands during task performance, children with ASD perform similarly to their TD peers, although it has been suggested that participants with ASD may use a different processing strategy than TD controls (e.g., faster processing style) (Pexman et al., 2011). Another study that imposed similarly minimal demands during irony comprehension task performance (utilizing computer-supported conversations to reduce social demands and forced-choice questions to reduce verbal demands) found participants with ASD as accurate as TD controls in comprehending irony and judging the speaker's intent (Glenwright & Agbayewa, 2012). In line with these findings, Au-Yeung, Kaakinen, Liversedge, and Benson (2015) examined irony comprehension among adults with ASD as compared to TD adults by recording eye movements. Adults with ASD performed as well as TD controls on comprehension accuracy for both ironic and non-ironic conditions. In addition, eye movement data showed that total reading time was longer in the ironic condition, as compared to the non-ironic condition, in both study groups. Thus, the results of studies of irony comprehension in ASD are inconclusive, with some showing poorer performance of children with ASD, despite language abilities similar to TD controls, and others demonstrating a similar ability to comprehend irony. This inconsistency may be linked to the complexity of the task utilized. When a task poses reduced demands or is accompanied by various clues, irony comprehension may be improved.

One of the major theories explaining the source of the observed difficulty in comprehending figurative language in ASD is the Theory of Mind (ToM) (Baron-Cohen, Leslie, & Frith, 1985). This theory claims that people with ASD have difficulty grasping the mental state of the other, making it difficult for them to understand social situations. During figurative language comprehension tasks, the listener must understand the speaker's mental state and infer his or her intentions accordingly. Hence, mentalization difficulties prevent the listener from seeing the pragmatic intentions behind figurative expressions. This theoretical framework has been supported by previous studies (Happé, 1995; Whyte, Nelson, & Scherf, 2014). For example, Happé (1995) found that children with ASD who passed first-order ToM tasks were able to comprehend metaphors, but not irony. In contrast, children with ASD who passed second-order ToM tasks were able to perform tasks that require understanding metaphor as well as irony. However, in later studies this association between ToM and figurative language comprehension was challenged (Adachi et al., 2006; Angeleri & Airenti, 2014; Huang, Oi, & Taguchi, 2015; Norbury, 2004; Whyte et al., 2014). It was shown that Japanese children with high-functioning ASD (HFASD) comprehend irony even though there is no evidence that they have second-order ToM abilities (Adachi et al., 2006). In a recent study, the comprehension of figurative language (i.e. metaphors, irony, sarcasm and indirect requests) was investigated in Taiwanese children with HFASD compared to control TD children matched for age, gender, and receptive language (Huang et al., 2015). No differences in the comprehension of all figurative language types were observed between the group who passed only a first-order ToM task and the group that passed the second-order order ToM task. Consistent with these findings, it has been suggested that ToM is necessary but not sufficient for idiom comprehension (Norbury, 2004). In view of these inconsistent findings, the current study aimed to further evaluate the association between ToM abilities and irony comprehension. In the present study, ToM abilities were assessed via both a second-order false belief task called the *"Ice-cream van"* (Perner & Wimmer, 1985), as well as the Hinting test (Corcoran, Mercer, & Frith, 1995). The latter evaluates the participant's mentalizing ability and comprehension of intentions in diverse daily situations (Turner-Brown, Perry, Dichter, Bodfish, & Penn, 2008).

Other scholars attribute the difficulty to comprehend figurative language in individuals with ASD to executive dysfunction. Executive functions (EF) include elements of mental flexibility, inhibition, planning, working memory and monitoring of operations. With regard to figurative language, difficulty in executive functions can make it difficult for an individual to shift from one meaning (e.g., the literal meaning) to another (e.g., the figurative interpretation) (Cummings, 2009; Landa & Goldberg, 2005). Indeed verbal ability and EF were correlated with proverbial understanding in children with HFASD but not in TD children (Tzuriel & Groman, 2017). Executive functions also contributed to the prediction of novel metaphor generation (Kasirer & Mashal, 2016) and novel metaphor understanding (Mashal and Kasirer, 2011) in children with ASD. According to Kasirer and Mashal (2014), among adults with ASD, comprehension of novel metaphors was best predicted by the Trail Making Test (TMT-B), a known neuropsychological assessment tool used to examine mental flexibility (Tombaugh, 2004). In the current study we examined the association between irony comprehension and executive functions using the Trail Making Test (TMT).

A different approach argues that individuals with ASD do not have a specific deficit in figurative language but rather a more

general language comprehension deficit (Gernsbacher & Pripas-Kapit, 2012). Several studies (Norbury, 2004; Whyte et al., 2014) have found a link between language abilities and idiom comprehension in individuals with ASD. A previous study (Saban-Bezalel & Mashal, 2015) investigated the comprehension and hemispheric processing of both irony and idioms in adults with ASD and TD adults. The results showed that when the two groups were matched on nonverbal IQ and age, with better (though nonsignificant) vocabulary in the TD group, the TD participants outperformed the participants with ASD in the irony and idiom tasks, and the two groups had different patterns of hemispheric processing. After careful matching of the two groups on vocabulary scores, however, the differences between the groups in idiom comprehension disappeared but remained significant in irony comprehension. These findings accord with a recent meta-analytic review (Kalandadze, Norbury, Narland, & Nass, 2016) that examined figurative language abilities in TD individuals and those with ASD. The researchers found that different subtypes of figurative language cannot be considered the same, as each of them makes use of different abilities. Moreover, the differences between participants with ASD and controls were less when the groups were matched on the basis of language abilities.

Despite the evidence documenting difficulties in figurative language comprehension in ASD individuals, there is no consensus among researchers about the cause of these difficulties. There is also controversy about how mentalizing abilities are linked to figurative language comprehension, and irony comprehension in particular. The present study thus had two main goals: 1) to investigate irony comprehension in individuals with ASD as compared to TD controls matched on expressive vocabulary, executive functions, and second-order ToM understanding, which have been found to contribute to the comprehension of figurative language; and 2) to examine the contribution of mentalizing ability to irony comprehension. We hypothesized that differences in irony comprehension would be lower when the groups were well-matched for language, EF and ToM. We also hypothesized that ToM abilities would contribute positively to irony comprehension.

2. Method

2.1. Participants

Forty-eight native Hebrew speakers, 26 of whom were participants with ASD (21 males) (age range 9–18.5) and 22 of whom were TD participants (17 males) (age range 7.5–15) were recruited for the study. A simple matching procedure was applied to ensure similar age range in the two groups. This led to the exclusion of eight participants: six participants above fifteen years old from the ASD group and two participants below ten years from the TD group. This yielded a group of 20 participants with ASD (M = 12.84, SD = 1.42), and 20 participants with TD (M = 11.96, SD = 1.46), with no age difference between the groups. For demographic characteristics, see Table 1.

All the participants in the ASD group had been previously diagnosed by an independent psychiatrist following DSM-IV-TR criteria (American Psychiatric Association, 2000). The clinical diagnoses included PDD-NOS (n = 11, 42%), autism (n = 10, 38%), and Asperger's syndrome (n = 5, 20%). In light of the new classification that now appears in the DSM-5 (American Psychiatric Association, 2013), we refer to the entire group as ASD. To confirm their diagnosis, the participants with ASD were also assessed with the Social Communication Questionnaire (Berument, Rutter, Lord, Pickles, & Bailey, 1999). The items on this questionnaire cover three areas of functioning: reciprocal social interaction, language and communication, and repetitive and stereotyped behavior. All the participants in the ASD group received a score above 15 on this questionnaire, further verifying the clinical diagnosis. Prior to their children's participation in the study, all the parents received an introductory letter that explained the experiment's purpose and method, and signed a consent form. The participant recruitment adhered to Bar Ilan University's institutional research guidelines, and the study was approved by the Israel Ministry of Education. See Table 1 for the demographic characteristics of the participants with ASD and the TD controls.

Table 1
Group comparison of participants' characteristics and performance of vocabulary, ToM, Hinting test, TMTs and literal strips.

	ASD (N = 20)		TD (N = 20)				
	Mean	*SD*	*Mean*	*SD*	*T(38)*	*p*	*Cohen's d*
Age	12.84	1.42	11.96	1.46	1.93	.06	0.61
Gender (males)	17		16		.17[!]	.68	–
Vocabulary	39.95	7.76	39.55	7.19	0.17	.87	0.05
ToM	0.80	0.41	0.85	0.37	−0.41	.69	0.13
Hinting test	13.30	3.23	17.15	1.63	−4.76***	< .001	1.50
TMT-A	51.80	20.21	53.95	26.60	−0.29	.78	0.09
TMT-B	117.30	31.27	106.30	31.27	0.74	.47	0.35
Literal comprehension[1]	91.00	18.89	98.00	6.16	−1.58	.12	0.50

*** $p < .001$.
[!] Results of $\chi^2(1)$ test.
[1] Mean across five repeats.

2.2. Stimuli

2.2.1. Screening tests

2.2.1.1. Social–cognitive tests. Theory of Mind Second-Order False-Belief Attribution Task "Ice-cream van" (Perner & Wimmer, 1985). This is an acted story task. Two characters (John and Mary) are in the park when an ice cream van arrives. John wants to buy ice cream but has no money. The ice cream vendor reassures John that he can go home to get the money because he will stay in the park all afternoon. John goes home, but in the meantime, the ice cream vendor changes his mind and decides to go to a school to sell his ice cream instead. Since Mary has remained at the park, she knows that the ice cream vendor has left and she also knows that he told John he would remain at the park. On the way to the school the ice cream van passes by John's house. John sees the vendor, talks to him and learns about his change of location. Since Mary has remained at the park, she doesn't know that John and the ice cream vendor have met. Mary goes to look for John at his house. John's mother tells Mary he has gone to buy ice cream.

Throughout the story, the participants answer control questions to make sure they understand the story (e.g., Does Mary know that the ice-cream man has talked to John?). All participants responded correctly to all the control questions, suggesting intact understanding of the story narrative. Only at the end of the story the participants answer a question related to second-order belief, namely, "Where does Mary think John went to buy ice cream?" The participant scores 1 point for the correct answer ("the park") and 0 for any other answer.

2.2.1.1.1. Hinting test. The Hinting test is used to measure ToM skills and comprehension of intentions (Turner-Brown et al., 2008). The task was originally designed by Corcoran et al. (1995) and was translated by a professional translator into Hebrew. The test consists of ten short stories describing interactions between two speakers. Each story ends with one speaker dropping a hint. The participant is asked what the speaker really meant to say. A correct answer describes the speaker's tacit intention. Each correct answer gives the participant two points. If the participant fails to give the correct response, more information is added. The participant is asked again what the speaker wants the other character to do. A correct response provides a score of one point. If the participant again fails to infer the intended meaning, a score of zero is given for the item. The maximum score on this test is 20. The items (the short stories) were read aloud to the participants with appropriate prosodic intonation (Corcoran et al., 1995). The Hinting test has been used in a variety of studies assessing ToM abilities and has good psychometric properties (e.g., Corcoran et al., 2001; Marjoram et al., 2005). An example of an item: Rebecca's birthday is approaching, so she says to her dad, "I love animals, especially dogs." *Question:* What does Rebecca really mean when she says this? (If the participant fails to answer s/he is given additional information) *Add:* Rebecca goes on to ask, "Is the pet shop open on my birthday?" *Question:* What does Rebecca want her dad to do?

2.2.1.1.1.1. Vocabulary

Participant vocabulary was tested with the vocabulary subtest of the Wechsler Intelligence Scale for Children (WISC-IV^{HEB}; Wechsler, 2003). In this task participants hear words and are asked to provide their definitions.

2.2.1.1.1.2. Executive functions

The Trail Making Test (TMT; Reitan and Davison, 1974) is a known neuropsychological test that is used to examine executive functioning and in particular mental flexibility (Tombaugh, 2004). This test has been linked to figurative language comprehension (Mashal and Kasirer, 2012). The test consists of two parts, A and B. In part A, participants are given a page with a series of numbers and are asked to connect them in ascending order. In part B, participants are asked to connect a series of numbers and letters in ascending order; this task requires the participant to alternate between numbers and letters. The results, which are based on the time it takes the participant to complete the task, are converted to Z-scores.

2.2.1.1.2. Irony comprehension task. The participants are given an irony task, which includes 15 comic strips, 10 with ironic interpretations and 5 with literal interpretations. The comic strips are presented in a pseudo-random order. Following each comic strip, the participants are asked an open question that assesses irony comprehension. Each correct answer receives one point (with a maximum of 10 points for the ironic comic strips and a maximum of 5 points for the literal ones). The task begins with two practice trials, one ironic and one literal. See Fig. 1 for an example of a comic strip with an open question.

2.2.1.1.3. Stimulus construction. Comic strips were taken from a pool of 84 passages, some with an ironic interpretation and some with a literal one (Saban-Bezalel & Mashal, 2015), and the final word of each strip gives the passage its literal or ironic meaning. We recruited 20 adult judges (aged 18–35) who were presented with the passages with the final word was missing and were asked to write the word that gives the passage its literal or ironic interpretation. The words used by at least 80% of the judges were chosen. To validate whether the passage was ironic or literal, the full passages (including the selected final words) were presented to 20 additional judges (aged 18–35). The judges were asked to indicate whether the passage was ironic, literal, or meaningless. The judges' agreement rates were 97% (SD = 0.06), 96% (SD = 0.81), and 92% (SD = 0.11), for the literal, the ironic and the meaningless endings, respectively. A set of 10 ironic and 5 literal strips were selected and constructed using the Toondoo website (http://www.toondoo.com). Finally, three additional judges (aged 18–25) were presented with these comic strips and asked to indicate whether each one was ironic or literal. The agreement rates were 100% and 97% (SD = 0.11) for the literal and the ironic strips, respectively. The demographic characteristics and the comprehension of the literal strips are presented in Table 1.

As can be seen in Table 1, except for Hinting test differences ($t(38) = -4.76, p < .001$), the two groups did not differ in gender distribution, age, vocabulary, ToM scores, executive functions (as assessed by the TMT), and literal comprehension of the comic strips. Note that groups differed moderately in age as p value was nearly significant ($p = .06$).

2.3. Procedure

The aim of the study was to investigate whether children and adolescents with ASD differ from their TD peers in their ability to

Fig. 1. An example of a comic strip.

comprehend irony when matched on language ability, second-order ToM, and executive functions. First, all the participants underwent the second- order ToM task. Then the vocabulary, Hinting, executive function, and the irony comprehension tasks were administered in random order. The participants were tested separately at school or at home. The study was conducted in one session.

3. Results

A preliminary comparison between the groups on accuracy scores (rates of successful answers) for the ironic strips revealed a significant difference, $t(38) = 1.84$, $p < .05$, d' = .58 (one-tail test) (ASD: $M = 79$, SD = 20.36; TD: $M = 89$, SD = 13.33). These results provide an initial indication of possible differences in understanding irony between the groups.

3.1. Correlations of vocabulary, hinting, ToM, and Trail Making Test scores with irony comprehension in each group separately

Before conducting the regression analyses, we report bivariate correlations in each group separately to allow evaluation of the strength of the links between the predictor variables (vocabulary, Hinting test, Trail) and irony comprehension. No significant correlation was found between vocabulary and irony score in the TD group ($r(18) = -.32$, $p = .17$), and in the ASD group ($r(18) = .12$, $p = .61$). A significant correlation was found between Hinting test score and irony comprehension in the ASD group ($r(18) = .49$, $p < .05$) but not in the TD group ($r(18) = .34$, $p = .13$). We used a Fisher z test to test the significance of the difference between these correlations. No significant difference was found, $z = .53$, $p = .29$. No significant correlations were found between Trail and irony comprehension scores in the TD group ($r(18) = -.37$, $p = .11$) and in the ASD group ($r(18) = -.13$, $p = .59$).

As groups were not perfectly matched on age ($p = .06$) we also calculated the correlations between irony comprehension and age. No significant correlations were found between age and irony comprehension scores in the TD group ($r(18) = -.12$, $p = .61$) and in the ASD group ($r(18) = .37$, $p = .11$).

3.2. A regression with logit link

To test our research hypotheses, we developed a two-level binary logistic model, in which level one data were single measurements of irony comprehension on a dichotomous scale, and level two data were participants' rates of correct responses across the ten repeats (of the ironic short stories) as dependent variables (HLM V.7.03: Heck, Thomas, & Tabata, 2012; Raudenbush, Bryk, & Congdon, 2013). Due to the relatively small sample (N = 40), we performed several independent models, each of which tested a different key explanatory variable. Table 2 shows the model results. In other words, due to the small sample size we were left with little degrees of freedom that did not enable us to perform a more complex analysis such as an inferential comparison between the predictors. We therefore performed a descriptive comparison between model effects. That is, we analyzed which of the predictors significantly predicted irony comprehension.

The overall indication of variation across participants is given in the intra-class correlation coefficient (ICC = .21) based on the unconditional model results (Model 1). The mean intercept (1.79) means that the overall expected probability of successful answers was 0.86 (corresponding to 84 percent of observed successful answers, $p_{expected} = 1 / (1 + EXP[-1.79])$. These probabilities varied randomly across participants (random variance was 0.87, $p < .001$). Next, we added the group effect at level two (presented in Model 2), the participant's level, which resulted in a significant difference ($b = 0.81$, $p < .05$), or an additional probability of ten percent to provide the correct answer for the TD group vs. the ASD group (p_{group}-$p_{expected} = 1/(1+EXP[-1.41-0.81])-1/(1+EXP[-1.41])$. Research variables (vocabulary, hinting, ToM, Trail) were added one at a time (Model 3a to Model 3d, respectively) to show

Table 2
Logistic model of correct answers by independent variables.

	Estimate	*SD / SE*	*df*	OR	95% CI
Unconditional Model 1					
Intercept	1.79***	0.19	39	5.97	[3.94, 9.05]
Random Effect, Variance	0.87***	0.93	χ^2=84.18***		
Main Effects Model 2					
Intercept	1.41***	0.26	38	4.11	[2.41, 7.01]
Group	0.81*	0.41	38	2.25	[1.00, 5.10]
Random Effect, Variance	0.81***	0.90	χ^2 = 77.09***	$\Delta\chi^2_{(1)}$ = 7.09*	
Main Effects Model 3a					
Intercept	1.42***	0.27	37	4.15	[2.37, 7.26]
Vocabulary	0.03	0.02	37	1.03	[0.99, 1.08]
Random Effect, Variance	0.82***	0.91	χ^2 = 75.13***	$\Delta\chi^2_{(1)}$ = 1.96	
Main Effects Model 3b					
Intercept	1.89***	0.28	37	6.59	[3.75, 11.59]
Hinting	0.21***	0.06	37	1.23	[1.10, 1.38]
Random Effect, Variance	0.61***	0.79	χ^2 = 64.03***	$\Delta\chi^2_{(1)}$ = 13.06***	
Main Effects Model 3c					
Intercept	1.43***	0.27	37	4.19	[2.41, 7.28]
ToM	0.37	0.40	37	1.45	[0.65, 3.25]
Random Effect, Variance	0.86***	0.93	χ^2 = 77.17***	$\Delta\chi^2_{(1)}$=-0.08	
Main Effects Model 3d					
Intercept	1.44***	0.28	37	4.23	[2.42, 7.39]
Trail	−0.28!	0.16	37	0.75	[0.55, 1.04]
Random Effect, Variance	0.81	0.90	χ^2 = 74.68***	$\Delta\chi^2_{(1)}$ = 2.41	

ICC = .19; *N* = 48 participants; within participant's repeats – 10 items of ironic short stories; Trail = averaged TMTs; note that in all model 3 repeats (a–d), group was included.
$p < .01$.
* $p < .05$.
*** $p < .001$.
! $p = .053$.

a positive effect of hinting ($b = .21$, $p < .001$) and negative effect of the Trail task ($b = -0.26$, $p < .10$), where the Trail score is the standardized average of the two TMTs. The reduction in the fit between the model and the observed data ($\Delta\chi^2$) indicates improvement in the explanatory power of the model due to the additional explanatory variables beyond the group effect.

A further interaction analysis (group*Independent Variable) was added to Models 3a to 3d, to estimate different performance effects on irony comprehension between the ASD and TD groups. The only significant interaction was found between group and Trail (Model 3d, interaction = −0.59, *SE* = 0.26, $p = .031$), where Trail had a negative effect on irony across participants with TD (b = -0.78, *SE* = 0.22, $p = .002$), but not across participants with ASD ($p = .491$). Thus, faster performance on Trail, i.e., higher mental flexibility, is linked to better irony comprehension in TD.

3.3. Analysis of subgroups matched on hinting test performance

Our analyses showed that when the two groups were matched on age and vocabulary scores (but differed on their performance in the Hinting test), the TD group outperformed the ASD group in irony comprehension ($M = 89$, $M = 79$, respectively). We also found a significant contribution of the Hinting test scores to the prediction of irony comprehension. Therefore, we adjusted group matching by the Hinting test scores and reanalyzed the data for these subgroups. Each participant with ASD was matched with a TD participant with the same or a close score on the 'Hinting test', forming a subgroup of 12 ASD participants and 12 TD participants selected for this analysis. The mean hinting scores on this ASD subgroup was $M = 15.33$, $SD = 2.30$; and $M = 16.25$, SD = 1.48 in the TD group. *t*-test revealed no significant irony comprehension difference between this ASD subgroup ($M = 83.33$, $SD = 24.62$) and the TD subgroup ($M = 85.83$, $SD = 13.79$), t(22) = 1.24, $p = .76$. See Fig. 2. This shows that when the two subgroups were matched on mentalizing ability as assessed by the Hinting test they show similar performance on the irony comprehension task. However, as sample size was reduced for this comparison, the probability of a Type II Error increased; the null hypothesis was not rejected, although rates of irony comprehension were similar to those of the full sample.

3.4. Descriptive statistics of the irony and the hinting scores according to ASD subtypes

In this final analysis, we returned to the initial pool of participants with ASD (N = 26). Our sample included nine participants with autism, twelve participants with PDD, and five participants with Asperger syndrome (AS). Although the DSM-IV criteria propose that all of these diagnoses fall within the autism diagnostic spectrum, we sought to explore their separate performance levels on the irony comprehension task as well as the Hinting test. The mean scores on the ironic comic strips were 81.12 ($SD = 12.70$), 63.34 ($SD = 28.40$), and 88.00 ($SD = 13.04$) in the autism, PDD, and AS subgroups, respectively. Mean performance on the Hinting test was

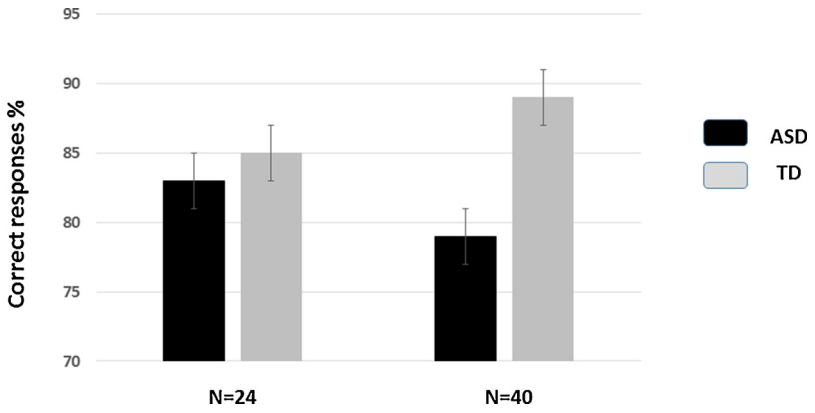

Fig. 2. Results of the irony comprehension test in the groups matched on vocabulary (N = 48) and on performance in the Hinting test (N = 30).

13.56 (*SD* = 2.79), 11.83 (*SD* = 3.15), 15.66 (*SD* = 2.97), in the autism, PDD, and AS subgroup, respectively. The participants with AS seem to score higher on both tasks, consistent with past studies showing differences in cognitive processing between these subgroups. Given the small sample sizes of these subtypes in our current study, this is a preliminary finding; future studies with larger numbers of participants are encouraged to replicate this analysis and further explore whether the delay in language development among those with autism could be a potential factor negatively impacting their irony comprehension.

4. Discussion

The present study investigated the ability of individuals with ASD to comprehend irony. Our findings indicate that, although the two groups were matched on age and scores from a second-order ToM task, as well as their level of vocabulary and executive functioning, children and adolescents with ASD understood fewer ironic comic strips than their TD peers. However, when the two groups were further matched on their mentalizing ability (as assessed by the Hinting test), the difference was no longer significant. Understanding of others' intentions (as measured by the Hinting test) and mental flexibility predicted irony comprehension beyond group and gender.

The first aim of the current study was to examine irony comprehension in children and adolescents with ASD as compared to TD controls matched on expressive vocabulary, mental flexibility, and a second-order ToM task, all of which have been found to contribute to figurative language comprehension (Angeleri & Airenti, 2014; Filippova & Astington, 2008; Gernsbacher & Pripas-Kapit, 2012; Happé, 1995; Kasirer & Mashal, 2014). We hypothesized that differences in irony comprehension would be lower when the groups were well-matched. Although the children with ASD in our study exhibited good irony comprehension (79% correct responses), nonetheless, despite the careful matching, the results revealed a significant difference between the two groups (89% vs. 79%). To the best of our knowledge, such group matching was not performed in previous studies that tested irony comprehension. Thus, although these abilities are essential for irony comprehension in ASD, they are not sufficient to fully explain the differences observed between the two study groups. Our results also show that the group difference disappeared after more careful matching in terms of Hinting scores, a measure of mentalizing ability. In a previous study that tested irony comprehension using the divided visual field paradigm (Saban-Bezalel & Mashal, 2015) adults with pervasive developmental disorder (PDD) also performed less accurately than did their TD peers. However, in contrast to the present study, this group difference did not disappear after adjusting group matching for vocabulary knowledge. Thus, whereas the PDD group did less well, even when matched for vocabulary, the current findings show that adjusting group matching according to mentalizing ability eliminates the difference in irony comprehension. In line with these findings, the results of the regression analysis also emphasize the importance of mentalizing ability over vocabulary knowledge, as Hinting test (but not vocabulary) scores predicted irony comprehension.

The second aim of the study was to examine the contribution of mentalizing ability to irony comprehension. We hypothesized that ToM abilities would contribute positively to irony comprehension. Indeed, we found that ToM, as assessed by the Hinting test, contributed to irony comprehension. The relationship between ToM and figurative language comprehension is in line with some previous studies (e.g., Happé, 1995) but is challenged by others (Adachi et al., 2006; Angeleri & Airenti, 2014; Huang et al., 2015; Norbury, 2004; Whyte et al., 2014). In the current study we used two different tasks to examine mentalizing abilities: one of second-order false beliefs and the other of comprehension of intention (Hinting test). This is the first time that these two tests have been used together to test ToM abilities in ASD. Indeed, most studies that examined ToM abilities in ASD used tests of first-order false beliefs, second-order false beliefs, or both (Adachi et al., 2006; Angeleri & Airenti, 2014; Happé, 1995; Huang et al., 2015; Norbury, 2004), without testing comprehension of intention. Other studies used different tasks (i.e., 'strange stories' and 'reading the mind in the eyes') (Whyte et al., 2014). Our findings reinforce the relationship between ToM abilities and irony comprehension, and emphasize the importance of using tests that relate to various aspects of mentalization when examining ToM abilities.

What could possibly be the difference between the two mentalizing tests (second-order false belief task and Hinting test)? Comprehension of others' mental states is a complex ability that includes various aspects such as belief, knowledge, perception, intention and emotion (Tahiroglu et al., 2014). Each of the tests we used focus on different mentalizing abilities. The Hinting test

focuses on the comprehension of intentions (Corcoran et al., 1995; Turner-Brown et al., 2008) and the ability to infer the intentions behind an indirect speech utterance. The participant is given limited context and has to perceive that the intention is different from what is said (Corcoran et al., 2001). While the Hinting test focuses on the comprehension of others' intentions in various everyday life situations, and is based on more complex components including rich scenarios and world knowledge, the second-order ToM task examines the comprehension of high-order beliefs. In the second-order false belief task (i.e.,"Ice-cream van"), the listener hears a story acted out and is then asked to predict the thought of one person based on the thoughts of another person in the context of conflicting scenarios. One scenario is knowledge of the van's location (i.e. "the van is in the park" vs. "the van is at the school") and the other is a conflict between what John knows (i.e., "John knows the van is at the school") and what Mary thinks John thinks (i.e., "John thinks the van is in the park") (Perner & Wimmer, 1985). The greater association observed between irony comprehension and Hinting test scores rather than the false belief task, may indicate that different aspects of mentalizing ability are addressed by each task, and this should be tested in future studies of figurative language comprehension.

Mental flexibility also contributed to irony comprehension. Mental flexibility may facilitate the listener's ability to shift between the literal meaning of the utterance and the contextual cues, and to infer the intention of the speaker. Interestingly, when we examined the interaction between group and mental flexibility, an effect on irony performance was observed in the TD but not in the ASD group. This finding is consistent with studies that found no relationship between executive functions and figurative language in ASD (Landa & Goldberg, 2005) and also in schizophrenia (Champagne-Lavau & Stip, 2010). Our findings therefore suggest that whereas mental flexibility (as assessed by the TMT) does not contribute to irony comprehension in ASD, mentalizing ability (Hinting test) does so for both ASD and TD participants. It should be noted that most of the studies that examined the relationship between EF and figurative language in ASD focused on metaphorical language (Kasirer & Mashal, 2016; Landa & Goldberg, 2005; Mashal and Kasirer, 2011; Tzuriel & Groman, 2017) and not on irony comprehension. Future studies are needed to test the different abilities that contribute to irony comprehension in ASD and in particular, the contribution of mentalizing abilities and EF.

In the current study, marginally significant differences in irony comprehension were found between children with ASD and their TD matched peers. Although this finding is in line with several studies (Mackay & Shaw, 2004; Wang et al., 2006), it is inconsistent with others (Colich et al., 2012; Pexman et al., 2011). The lack of agreement between the different studies is not surprising in light of differences in the methods used. Differences are found in the way the task was presented (e.g., cartoon, puppet show, computer-mediated conversation, illustration), the requirements for performing the task (e.g., answer a comprehension question verbally, forced-choice questions which reduce the linguistic demands), and the range of clues given to the participant (e.g., visual, prosodic) (Colich et al., 2012; Glenwright & Agbayewa, 2012; Mackay & Shaw, 2004; Pexman et al., 2011; Wang et al., 2006). Apparently, these differences created variability in the results, as performance by individuals with ASD is likely to be poor when the task is more demanding. Hence, our results support previous studies (Mackay & Shaw, 2004; Wang et al., 2006) that used fewer clues and a more demanding task, and found group differences in performance. Perhaps the comic strips we used focused on the visual aspect of the scenario and were pleasant to read but did not contribute enough to the comprehension of irony.

The present study has several limitations. First, the participants with ASD had all been diagnosed according to DSM-4 or DSM- 5, but the study lacked more robust diagnostic measures, such as the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) or the Autism Diagnostic Observation Schedule (ADOS-G; Lord et al., 2000). Such diagnoses could further shed light on the association between autistic traits and irony comprehension. Second, language abilities were tested solely via expressive vocabulary (not verbal IQ) and literal comprehension abilities. Our correlation analysis found no significant correlation between irony comprehension and vocabulary among the participants. We suggest that instruments that test broader language skills such as syntax, or perhaps full IQ assessment, should also be used in future studies. Another limitation of the current study is the relatively small sample size. Therefore, conclusions should be considered with caution. Due to the small sample size and the little degrees of freedom left, we could not test the effect size within one integrative model, meaning, we could not perform an inferential comparison between the predictors. Instead, we assigned a separated independent model for each predictor (vocabulary, hinting, ToM, Trail) that take into account the multiple repeats (the variability between the items). Thus, we performed a descriptive comparison between model effects and we reported which of the predictors significantly predicted irony comprehension. A further study with larger sample is necessary to compare effects in the assessment of irony comprehension. It should be noted however, that despite the small sample Hinting is a highly significant predictor of irony comprehension. We also acknowledge that the presentation of the ironic utterances in our study does not fully simulate ironic situations in everyday life, where irony comprehension requires rapid processing (i.e., identifying intentions in a particular context at the appropriate time). The current study did not set a time limit for the participants' processing. Finally, the current study tested only the accuracy rate of irony comprehension, but this design cannot tell whether there are any temporal differences between the two groups. Utilizing the current study methodology but examining eye movement in addition would allow investigation of whether the irony comprehension in ASD occurs on-line or whether there is temporal processing or detection time differences in irony comprehension.

To summarize, the present study investigated irony comprehension among children and adolescents with ASD as compared with their language-matched TD peers. Although the ASD group demonstrated adequate irony comprehension ability, the TD group outperformed the ASD group. When we reanalyzed the results with two subgroups adjusted according their Hinting test scores, the difference in irony comprehension disappeared. In addition, Hinting test scores and executive functions contributed to irony comprehension. The study shows that in addition to intact language abilities, different aspects of mentalizing abilities should also be considered in future studies assessing irony comprehension. The current results emphasize the importance of using sensitive instruments for evaluating mentalizing abilities.

References

Adachi, T., Hirabayashi, S., Shiota, M., Suzuki, S., Wakamiya, E., Kitayama, S., et al. (2006). The study of situational recognition of attention deficit/hyperactivity disorders, Asperger's disorder and high functioning autism with the metaphor and sarcasm scenario test (MSST). *No to Hattatsu Brain and Development, 38*, 177–181.

Aguert, M., Laval, V., Gauducheau, N., Atifi, H., & Marcoccia, M. (2016). Producing irony in adolescence: A comparison between face-to-face and computer-mediated communication. *Psychology of Language and Communication,* 199–218.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: Text revision* (4th ed. rev.). Washington, DC.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM V)* (5th ed.). Washington, D.C: American Psychiatric Association.

Angeleri, R., & Airenti, A. (2014). The development of joke and irony understanding: a study with 3- to 6-year-old children. *Canadian Journal of Experimental Psychology, 68*, 133–146.

Au-Yeung, S. K., Kaakinen, J. K., Liversedge, S. P., & Benson, V. (2015). Processing of written irony in autism Spectrum disorder: An eye-movement study. *Autism Research, 8*(6), 749–760. https://doi.org/10.1002/aur.1490.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition, 21*, 37–46.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *The British Journal of Psychiatry, 175*(444), 451. https://doi.org/10.1192/bjp.175.5.444.

Champagne-Lavau, M., & Stip, E. (2010). Pragmatic and executive dysfunction in schizophrenia. *Journal of Neurolinguistics, 23*, 285–296. https://doi.org/10.1016/j.jneuroling.2009.08.009.

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology General, 113*, 121–126.

Colich, N. L., Wang, A. T., Rudie, J. D., Hermandez, L. M., Bookheimer, S. Y., & Dapretto, M. (2012). Atypical neural processing of ironic and sincere remarks in children and adolescents with autism spectrum disorders. *Metaphor and Symbol, 27*(1), 70–92.

Corcoran, R., et al. (2001). Theory of mind and schizophrenia. In D. L. Penn (Ed.). *Social cognition and schizophrenia*Washington, DC: American Psychological Association pp. 149–174.

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: Investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research, 17*, 5–13.

Cummings, L. (Ed.). (2009). *Clinical pragmatics*. Cambridge: Cambridge University Press.

Dews, S., & Winner, E. (1995). Muting the meaning: A social function of irony. *Metaphor and Symbolic Activity, 10*, 3–19.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning in discourse irony understanding. *Child Development, 79*, 126–138.

Filippova, E., & Astington, J. W. (2010). Children's understanding of social-cognitive and social-communicative aspects of discourse irony. *Child Development, 81*(3), 913–928.

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*, 93–105.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol, 15*, 5–27.

Glenwright, M., & Agbayewa, A. S. (2012). Older children and adolescents with high-functioning autism spectrum disorders can comprehend verbal irony in computer-mediated communication. *Research in Autism Spectrum Disorders, 6*, 628–638.

Happé, F. G. E. (1995). Understanding minds and metaphors: Insight from the study of figurative language in autism. *Metaphor and Symbolic Activity, 10*, 275–295.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2012). *Multilevel modeling of categorical outcomes with SPSS*. London: Routledge.

Huang, S. F., Oi, M., & Taguchi, A. (2015). Comprehension of figurative language in Taiwanese children with autism: The role of theory of mind and receptive vocabulary. *Clinical Linguistics & Phonetics, 29*, 764–775. https://doi.org/10.3109/02699206.2015.1027833.

Kalandadze, T., Norbury, C., Narland, T., & Nass, K. B. B. (2016). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism,* 1–19.

Kasirer, A., & Mashal, N. (2014). Verbal creativity in autism: Comprehension and generation of metaphoric language in high functioning autism spectrum disorder and typical development. *Frontiers in Human Neuroscience, 8*, 818.

Kasirer, A., & Mashal, N. (2016). Comprehension and generation of metaphors by children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 32*, 53–63.

Landa, R. J., & Goldberg, M. C. (2005). Language, social, and executive functions in high functioning autism: A continuum of performance. *Journal of Autism and Developmental Disorders, 35*(5), 557–573.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview–revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659–685.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., 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*, 205–223.

Mackay, G., & Shaw, A. (2004). A comparative study of Figurative language in children with autistic spectrum disorders. *Child Language Teaching and Therapy,* 13–32.

Marjoram, D., Gardner, C., Burns, J., Miller, P., Lawrie, S., & Johnstone, E. (2005). Symptomatology and social inference: A theory of mind study of schizophrenia and psychotic affective disorder. *Cognitive Neuropsychiatry, 10*, 347–359.

Martin, I., & Mcdonald, S. (2004). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language, 85*, 451–466.

Mashal, N., & Kasirer, A. (2011). Thinking maps enhance metaphoric competence in children with autism and learning disabilities. *Research in Developmental Disabilities, 32*, 2045–2054.

Mashal, N., & Kasirer, A. (2012). The relationship between visual metaphor comprehension and recognition of similarities in children with learning disabilities. *Research in Developmental Disabilities, 33*(6), 1741–1748.

Norbury, C. F. (2004). Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech Language and Hearing Research, 47*, 1179–1193.

Perner, J., & Wimmer, H. (1985). John thinks that mary thinks that attribution of second-order beliefs by 5- to 10-year-old. *Children. Journal of Experimental child Psychology, 39*, 437–471.

Pexman, P. M., & Glenwright, M. (2007). How do typically developing children grasp the meaning of verbal irony? *Journal of Neurolinguistics, 20*(2), 178–196.

Pexman, P. M., Rostad, K. R., McMorris, C. A., Climie, E. A., Stowkowy, J., & Glenwright, M. R. (2011). Processing of ironic language in children with high- functioning autism spectrum disorder. *Journal of Autism and Developmental Disorders, 41*, 1097–1112.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2013). *HLM 7.01 for Windows [Computer software]*. Skokie, IL: Scientific Software International, Inc.

Reitan, R. M., & Davison, L. A. (Eds.). (1974). *Clinical neuropsychology: Current status and applications*. Washington, DC: Winston.

Roberts, R. M., & Kreuz, R. J. (1994). Why do people use figurative language? *Psychological Science, 5*, 159–163.

Saban-Bezalel, R., & Mashal, N. (2015). Hemispheric processing of idioms and irony in adults with and without pervasive developmental disorder. *Journal of Autism and Developmental Disorders, 45*, 3496–3508.

Tahiroglu, D., Moses, L. J., Carlson, S. M., Mahy, C. E. V., Olofson, E. L., & Sabbagh, M. A. (2014). The children's social understanding scale: Construction and validation of a parent-report measure for assessing individual differences in children's theories of mind. *Developmental Psychology, 11*, 2485–2497. https://doi.org/10.1037/a0037914.

Tombaugh, T. N. (2004). Trail making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology: the Official Journal of the National Academy of Neuropsychologists, 19*, 203–214.

Turner-Brown, L. M., Perry, T. D., Dichter, G. S., Bodfish, J. W., & Penn, D. L. (2008). Brief report: Feasibility of social cognition and interaction training for adults with high functioning autism. *Journal of Autism and Developmental Disorders, 38*, 1777–1784. https://doi.org/10.1007/s10803-008-0545-y.

Tzuriel, D., & Groman, T. (2017). Dynamic assessment of figurative language of children in the autistic spectrum: The relation to some cognitive and language aspects. *Journal of Cognitive Education and Psychology, 16*, 38–63.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain, 129*, 932–943.

Wechsler, D. (2003). *Wechsler intelligence scale for children: WISC-IV* (4th ed.). San Antonio, TX: Psychological Corporation.

Whalen, J., Pexman, P. M., Gill, A. J., & Scott, N. (2013). Verbal irony use in personal blogs. *Behaviour & Information Technology, 32*, 560–569.

Whyte, E. M., Nelson, K. E., & Scherf, K. S. (2014). Idiom, syntax, and advanced theory of mind abilities in children with autism spectrum disorders. *Journal of Speech Language and Hearing Research, 57*(1), 120–130.