Further Development in Social Reasoning Revealed in Discourse Irony Understanding

Eva Filippova and Janet Wilde Astington

University of Toronto

This study describes the development of social reasoning in school-age children. An irony task is used to assess 5-, 7-, and 9-year-olds’ (N = 72) and adults’ (N = 24) recursive understanding of others’ minds. Guttman scale analysis demonstrates that in order to understand a speaker’s communicative intention, a child needs to recognize the speaker’s belief, the detection of which depends on the ability to identify the discrepancy between the intended and the expressed meaning. Only children who understand these aspects of mind are able to reflect on the speaker’s attitude. Theory of mind and language ability make unique contributions to children’s interpretation of irony over and above the impact of age and memory, but attunement to expressive prosody does not.

Theory-of-mind (ToM) research investigates children’s ability to explain people’s behavior by reasoning about their mental states. Over the past two decades, researchers have focused on preschool children’s acquisition of mental-state concepts, such as desire, belief, and intention. By 3 years of age, children recognize that actions and preferences are determined by desires (Repacholi & Gopnik, 1997; Wellman & Woolley, 1990). By 5 years, they recognize that a person may hold a false belief about a state of affairs in the world, and they can use this knowledge to correctly explain or predict the person’s action (Wellman, Cross, & Watson, 2001). However, in order to explain people’s behavior, children need to reason about interpersonal interactions, not just about a person’s actions in the physical world. Importantly, much of our everyday social reasoning is concerned with dyadic communication and relationships, in which the actual or possible mental states of one member of the dyad must be seen as part of the intentional mental states of the other member.

The development of these more advanced aspects of social reasoning is underinvestigated in the theory-of-mind literature, perhaps leaving an erroneous impression that ToM is mastered by the end of the preschool years. However, only in the early school years do children acquire an understanding of the doubly embedded representations that are required to keep track of dyadic relationships. Perner and Wimmer (1985) showed that by 7–8 years of age, children are able to represent and reason from second-order beliefs (e.g., X believes that Y believes that p). That is, they understand that not only do people have beliefs (and false beliefs) about the world but they also have beliefs about the content of others’ minds (e.g., about others’ beliefs), and similarly, these too may be different or wrong. More recently, children 6 years of age showed this understanding on a simplified task (Astoning, Pelletier, & Homer, 2002). In addition, other tasks have been developed to assess children’s ability to deal with higher order representations involving intentions and emotions, as well as beliefs, for example, Happé’s (1994) “strange stories” and stories involving a “faux pas” (Banerjee, 2000; Baron-Cohen, O’Riordan, Stone, Jones, & Plaisted, 1999).

Understanding higher order representations of belief, intention, and emotion is required for the more mature understanding and use of complex language that develop during the school years, as in indirect speech, for example. Therefore, an investigation of children’s ability to comprehend indirect speech acts has the potential to illuminate the development of advanced social reasoning. A few studies have examined school-age children’s ability to distinguish different speech acts in relation to their theory-of-mind skills. For example, in order to differentiate between a joke and a lie, children need to detect
whether the speaker knows what the listener knows; that is, they need to be aware of the speaker’s second-order belief (Leekam, 1991) or second-order ignorance (Sullivan, Winner, & Hopfield, 1995).

Successful communication is contingent on inferring correctly what others mean by what they say, but in some communicative exchanges, such as jokes, the expressed and intended meanings may diverge. Happé (1993) demonstrated that an advanced level of mental-state attribution is necessary for correct interpretation of some communicative exchanges over others. Whereas first-order ToM is sufficient for understanding metaphors, an attribution of second-order representations, that is the ability to reflect on the speaker’s thought about an attributed thought, is needed for correct interpretation of discourse irony (Happé, 1993). However, the literature on children’s understanding of irony is somewhat unclear.

Irony Understanding

The use of irony entails indirectly conveyed beliefs and attitudes that constitute the speaker’s intended meaning. The listener must substitute the intended or implied meaning for the literal one expressed in the utterance. Irony achieves relevance by informing the addressee of the fact that the speaker has in mind a representation of something and, importantly, has a certain attitude toward it (Sperber & Wilson, 1995). Sperber and Wilson argue that irony invariably entails the implicit expression of an attitude (via paralinguistic and extralinguistic channels). Furthermore, the relevance of such an utterance depends on the information it conveys about the speaker’s attitude to the opinion expressed. An understanding of irony also requires recognition of the speaker’s motivation for the meaning substitution; for example, the use of “jocularity” is motivated by the speaker’s desire to express group solidarity (Gibbs, 2000).

On the surface, ironic statements frequently express false propositions. Although children are quite skilled communicators from fairly early in their lives, they are surprisingly unsuccessful in inferring the meaning behind false utterances until they enter school. Young children tend to interpret deliberately false utterances as sincere. When the speaker’s statement is incongruent with the situational facts, young children often discount the facts and interpret the utterance as wrong but sincere (Demorest et al., 1983). In one study (Demorest, Meyer, Phelps, Gardner, & Winner, 1984), 6-year-old children interpreted sarcastic remarks as sincere or mistaken about one third of the time and as deceptive—white lies—about half the time. The main distinction between deliberately false remarks in irony and in deception rests not only on a correct attribution of second-order beliefs (Winner, Brownell, Happe, Blum, & Pincus, 1998) but also on one’s understanding of the speaker’s communicative purpose or intention concerning the listener’s belief (Leekam, 1991; Winner & Leekam, 1991). Although 6-year-olds may recognize the nonliteral nature of irony, they fail to recognize speakers’ rationale for using such untrue statements, that is, their intention to mislead in the case of deception or the listener’s expected recognition of the falsity of the statement in the case of irony. Specifically, the ability to judge the literalness of a statement has been documented to precede the ability to interpret the speaker’s mind (Ackerman, 1983; Hancock, Dunham, & Purdy, 2000).

There is a proposed developmental sequence of understanding nonliteral language (Ackerman, 1981; Demorest et al., 1984) as follows: At 6 years of age, children interpret intentionally false utterances as sincere. Because they assume that the speaker’s belief and communicative intent are in accord with the statement, they rely on the statement as evidence of the speaker’s belief and intent, disregarding the story facts, intonation, or gestures. At around 9 years of age, children begin to understand deliberate falsehoods and make mature interpretations of inconsistencies between ironic utterances and contextual facts, but they are not fully aware of the intentional use of such falsehoods. Demorest et al. (1984) demonstrated that 9-year-olds continue to interpret lies and sarcasm as deceptive, taking account of the story facts and the speaker’s behavior when these are incongruent with the utterance but relying on the statement as evidence of the speaker’s communicative intent. Only after the early elementary school years can children understand the nuanced usage of nonliteral language by recognizing the purpose of the speaker’s statement. Only then are they able to distinguish between deceptions and irony and weigh the speaker’s behavior and the content of the statement to judge the speaker’s purpose (Demorest et al., 1984). Findings of Demorest et al. (1983, 1984) suggest that children as old as 13 years of age do not reliably distinguish irony from deception.

However, other investigators (Andrews, Rosenblatt, Malkus, Gardner, & Winner, 1986; Dew & et al., 1996; Sullivan et al., 1995; Winner, 1988; Winner & Leekam, 1991), using different methodologies, credit much younger children with the ability to interpret irony.
Such studies indicate that at least some of the components of verbal irony are grasped by the age of 6 years (e.g., the ability to detect discrepancy between sentence meaning and speaker meaning and to infer muting or humor function of irony). These inconsistent findings are due to the fact that the various studies investigate different components of irony, such as detection of meaning, awareness of the speaker’s emotions, beliefs, intentions, motivations, attitudes, or the function of irony. Little is known about the progression of children’s understanding of these different components that constitute an ironic utterance.

There is also disagreement over the extent to which children use prosodic cues to interpret ironic utterances. Prosody helps to convey the speaker’s emotion that may be a guide to utterance interpretation. Some researchers claim that children’s sensitivity to variations of acoustic parameters, such as pitch, loudness, and duration, contributes to their ability to understand irony (Capelli, Nakagawa, & Madden, 1990; de Groot, Kaplan, Rosenblatt, Dews, & Winner, 1995; Dews et al., 1996; Milosky & Ford, 1997). However, other researchers do not find these effects (Ackerman, 1983; Sullivan et al., 1995; Winner & Leekam, 1991).

The Present Study

The purpose of the present study is to examine advanced social reasoning, as revealed through children’s comprehension of irony. The first aim is to investigate the developmental sequence of understanding discourse irony by examining the different components that constitute an ironic utterance. The second aim is to investigate the relation between children’s understanding of discourse irony and their language skills, advanced ToM ability, and sensitivity to prosodic cues.

The various components of understanding others’ minds (i.e., understanding meaning, belief, communicative intention, and attitude) are separated in a newly developed irony task. Such a separation allows for documentation of a hypothesized acquisition sequence of the individual social-cognitive components of understanding a speaker’s mind. Children were expected first to reject the literal meaning of the utterance before they could detect the speaker’s belief and intention (Ackerman, 1983; Hancock et al., 2000). We further hypothesized that children’s detection of the speaker’s belief would precede detection of the speaker’s communicative intention (Carpendale & Chandler, 1996; Sokol & Chandler, 2003). Last, children were expected to identify the speaker’s attitude (Winner & Leekam, 1991).

Children’s performance on the newly developed irony task was examined in relation to their performance on ToM tasks assessing higher order understanding of mental states. Children’s understanding of discourse irony will also depend on their linguistic skills to some extent, in part because of the known close relation between language and ToM (Milligan, Astington, & Dack, 2007). Therefore, while controlling for general cognitive skills assessed by a memory measure, the study examined the extent of the contribution of children’s general language ability to their understanding of the ironist’s mind. Once age and memory are accounted for, we expected children’s scores on a general language measure and advanced ToM tasks to significantly predict their performance on the irony task. In addition, we also anticipated that children’s sensitivity to prosodic cues would independently predict their success on the irony task.

Method

Participants

Seventy-two children participated in the study: Twenty-four participants were classified as 5-year-olds (senior kindergartners; 11 boys and 13 girls; age range = 5.2 – 6.2, M age = 5.8 years), 24 participants as 7-year-olds (Grade 2 students; 13 boys and 11 girls; age range = 7.2 – 8.1, M age = 7.7 years), and 24 participants as 9-year-olds (Grade 4 students; 11 boys and 13 girls; age range = 9.1 – 10.1, M age = 9.7 years). The children came from primarily Caucasian, middle-to-upper-middle-income areas of a major Canadian city. Based on strict exclusion criteria, the sample was linguistically homogeneous, consisting of only monolingual English speakers. In addition, 24 adult volunteers (10 men and 14 women; age range = 19 – 56, M age = 37 years) took part in the study. All adult participants were native speakers of English.

Materials

Irony task. Eight different stories provided a measure of each participant’s ability to represent others’ minds. All eight scenarios were parallel in design: A child character was the recipient of an ironic comment by a fellow character. The stories were accompanied by simple illustrations of two story characters. In the illustrations, the character making an ironic comment was depicted from behind, facing the addressee, so that the mediation of interpretative cues through facial expressions was eliminated. To
assess the social-cognitive processes involved in interpreting the social scenarios and the speaker’s mind, each story was followed by a set of questions designed to quantify the individual components involved in complex reasoning about others’ minds. After the initial comprehension questions, each story was followed by questions assessing the detection of the meaning (context/utterance discrepancy), ToM (the speaker’s first-order belief and third-order intention), and motivation of the speaker (justification of why the speaker would say what he or she said; see Appendix A for example of a story and subsequent questions). All eight stories depicted situations familiar to children and represented scenarios in which children often participate or interact with parent, sibling, or friend, for example, helping in the kitchen, doing homework, shopping in grocery store, and eating in restaurant.

Advanced ToM tasks. Three kinds of advanced ToM tasks were used to test children’s understanding of second-order beliefs, second-order intentions, motivations, and attitudes (see Appendix B for examples of the tasks). Two second-order false belief stories (Astington et al., 2002), accompanied by illustrations, examined children’s awareness of a character’s knowledge of another character’s false belief. Two “strange stories” (Happe´, 1994), also accompanied by illustrations, examined children’s awareness of a character’s motivations to say or do something on the basis of his or her knowledge of another character’s belief about the event described in the story. Two “faux pas stories” (Banerjee, 2000), administered with the use of four props (three story characters and a target object), examined children’s awareness of a character’s knowledge of another character’s intentions, feelings, and motivations. Each task required either an explanation of a character’s behavior or a justification for a child’s attribution of a mental state to the story character.

Language. The Peabody Picture Vocabulary Test III (PPVT–III; Dunn & Dunn, 1997), a measure of receptive vocabulary, was used to assess children’s linguistic competence. It is a standardized test that correlates highly with other linguistic and cognitive measures.

Memory. The numbers subtest of the Children’s Memory Scale (CMS; Cohen, 1997) was used as a short-term memory measure. The task is based on repeating after the experimenter a series of single-digit sequences of increasing length (a) in the same order in which they were spoken (forward numbers) and (b) in the opposite order to that in which they were spoken (backward numbers).

Prosody task. A measure of children’s ability to detect emotions communicated through intonation was used to determine children’s sensitivity to prosodic cues. The stimuli were modeled on the standardized Florida Affect Battery (Bowers, Blonder, & Heilman, 1991) and were generated in Glenn Schellenberg’s laboratory (G. Schellenberg, personal communication, September 26, 2002). The task consists of 20 short digitally recorded utterances in a foreign language and is administered individually on a computer. The task is to match an emotion (happy, sad, angry, or afraid) depicted in one of four pictures on a computer screen with the emotion expressed in an utterance heard through a set of speakers on the basis of paralinguistic cues (acoustic features of speech) supplied by the voice of the person speaking in a foreign language.

Design and Procedure

Each child was interviewed individually in two sessions on two separate days in a testing room on the school premises. The procedure was identical for each participant.

Stories of both the irony and the advanced ToM tasks were read to children by the experimenter. During the administration of the irony task, the experimenter used the same stressed intonation for the ironic utterances across all the participants. Thus, all children received the same prosodic cues. Children also listened to six advanced ToM stories, two for each of the three kinds of tasks, and were questioned about the story characters’ thoughts and actions. Administration of these stories alternated with that of the irony stories. Passing comprehension questions in both types of stories was a prerequisite for inclusion in the data analysis. If a child did not respond correctly to either of the control questions, the stories were reread and the comprehension questions were asked once more. This was done to ensure that children remembered and understood the key points of each story.

The PPVT–III and the numbers subtest of the CMS task were administered according to standard procedures. In the prosody task, children listened to sentences spoken in Hebrew (it was established at the outset that they did not understand the language) and were asked to judge how a person uttering a sentence feels by pointing to one of four pictures on a computer screen depicting a happy person, a sad person, an angry person, or a scared person. The order of the stories within a task type was completely counterbalanced. Each session lasted for about 40 min.

Adult volunteers were tested on the irony task only. They were asked to fill out four questionnaires on the basis of the short stories that preceded them.
The stress in the ironic comments was marked. The protocols were identical to those used with children.

**Scoring and Coding**

In the irony task, a score was obtained from participants’ answers to the Meaning, Belief, Intention, and Motivation/Attitude questions on each story to determine the level of cognitive complexity of this task, as compared to the advanced ToM measures. The yes–no questions investigating the meaning, belief, and intention of the speaker were scored as either correct (1) or incorrect (0). Answers to the open-ended Motivation/Attitude question were coded on a 4-point scale, reflecting levels of interpersonal understanding based on increasing cognitive complexity of the answers, which were identified from the patterns of adults’ responses: Inappropriate justifications or two consecutive “I don’t know” answers received a score of 0; a score of 1 was given for responses reflecting simple surface-level justifications (i.e., relevant story facts, reference to states or feelings of the speaker) or those reflecting learned conventional answers or clichés; a score of 2 was given for a reference to information that has an implication for another person’s behavior, rather than mental state; a score of 3 was given to responses identifying the speaker’s intention to affect another person’s mental state; and a score of 4 was given for a reference made to a speaker’s attitude (whether positive or negative) toward the situation or to the pragmatic function of irony that goes beyond the identification of the speaker’s second-order intention (i.e., mention of what the speaker communicated to the listener about his or her attitude toward the situation; see Appendix C for the detailed coding scheme and examples of coded answers).

Based on the coding procedures used in the theory-of-mind literature (e.g., Astington et al., 2002; Happé, 1994), for each story a single score was given for responses to each question, whereby participants were credited for their best answer by being given the highest possible score. The responses of every participant on every story of the task were coded by a second independent coder naïve to the hypotheses being tested and blind to the ages of the participants. Inter-rater reliability ranged from 92% to 100% agreement for the individual stories (Cohen’s kappa ranged from .82 to 1). All discrepancies were resolved by discussion to a final resolution accepted by both coders.

The coding of justifications in the advanced ToM measures followed the coding schemes used in the original published studies (Astington et al., 2002; Banerjee, 2000; Happé, 1994). The coders’ agreement on the justifications ranged from 91% to 99% for the individual stories (Cohen’s kappa ranged from .78 to .98). All the discrepancies were resolved through a discussion in which both coders agreed to the final resolution. However, because some coding categories were represented only rarely in the children’s answers, the answers to the explanation question of each story were rescored on a scale from 0 to 2, corresponding to Happé’s (1994) coding scheme: A score of 0 was assigned for inappropriate justifications (e.g., irrelevant story facts) or two consecutive “I don’t know” answers; a score of 1 was assigned for an appropriate reference to a story fact that had relevance for the second-order reasoning; and a score of 2 was given for answers reflecting the second-order reasoning, that is, for mentioning of information that has an implication for one character’s mental state embedded within another character’s mental state (e.g., because Jim knows Simon is a liar; because Jim thought Simon was not telling the truth).

The scoring of PPVT – III and the numbers subtest of the CMS followed the standard scoring procedures. The software for the prosody task recorded the answers on 4 trial items and 16 test items. On this task, each child received a score based on the number of correctly matched emotions on the 16 test items.

**Results**

Results from a set of preliminary repeated measures analyses of variance (ANOVAs) provided evidence for the predicted age-related improvement in children’s understanding of an ironist’s mind. These analyses revealed that 5-year-olds lag consistently behind other age groups in performance on the four questions (Meaning, Belief, Intention, and Motivation/Attitude) and that adults’ performance exceeded that of all other age groups tested (Filippova, 2005).

A three-way ANOVA was conducted on the composite Irony scores (obtained as detailed in the next section), summed across the stories for each participant, with age (4), sex (2), and order (12) as the between-subject factors. The results revealed no main effects of sex or order and no interaction. However, the ANOVA revealed a significant age effect, $F(3, 92) = 18.86, p < .001$, partial $\eta^2 = .38$ (Figure 1). The Scheffé post hoc test showed significant differences between 5- and 7-year-olds ($p = .005$), between 5- and 9-year-olds ($p = .001$), between 5-year-olds and adults ($p < .001$), between adults and 7-year-olds ($p = .005$), and between adults and 9-year-olds ($p = .024$). Figure 1 illustrates that the overall means for participants’
scores on the irony task follow a clear developmental pattern. The differences in the group means are highly significant across the ages, with the exception of that between 7- and 9-year-olds.

**Scaling of Irony Social-Cognitive Components**

A developmental scale of social-cognitive components of the irony task was sought. Four variables were taken into account—Meaning, Belief, Intention, and Motivation/Attitude. Because the first three variables were dichotomous (i.e., yes vs. no), we also treated the scores on the Motivation/Attitude variable in a dichotomous fashion, whereby answers reflecting the speaker’s attitude (i.e., Score 4), observed as the most frequent justification in adults’ responses to the question, were treated differently from scores of 0–3, reflecting a motivation without an attitudinal component. The original score of 4 indicated an understanding that went beyond that reflected in the other three scaled variables (i.e., Meaning, Belief, and Intention; see Appendix C). (The means of the original scores 0–4 on the Motivation/Attitude question across the eight irony scenarios are as follows: 5-year-olds, $M = 1.33$; 7-year-olds, $M = 1.78$; 9-year-olds, $M = 2.33$; adults, $M = 3.02$.)

An Irony score based on the four irony social-cognitive variables was created for every irony story for each participant in a manner described in the upper part of Table 1. A new four-item scale was imposed on participants’ performance on the irony social-cognitive variables. According to Guttman (1944, 1950), scales that rank items in order of difficulty provide patterns of responses, according to which correct responses to an item imply correct responses to all previous items. The Guttman scalogram depicted in Table 1 provides an estimation of reproducibility of 5 identified patterns of the 16 potentially occurring patterns and identifies the extent to which responses to individual items fit the ideal pattern. Digression from perfection is measured by a coefficient of reproducibility, “the empirical relative frequency with which the values of the attributes correspond to the proper intervals of a quantitative variable” (Guttman, 1944, p. 140). According to Guttman, 85% correct scales are efficient approximations to perfect scales. To illustrate, Pattern 3 assumes that a child responded correctly to the meaning question of a given story (Meaning +) and to the belief question (Belief +) but responded incorrectly to the intention question (Intention $-$) and scored less than 4 on the Motivation/Attitude question for the same story (Attitude $-$).

The lower part of Table 1 summarizes the mean frequencies of occurrence of each pattern across the eight stories for each age group. In other words, the numbers represent means of patterns identified for each age group across all eight irony scenarios. As shown, a high proportion of responses fit the four-item Guttman scale exactly. Namely, 84.4% of the total number of responses (81 of 96) fit the scale. However, if the scalogram analysis excludes the youngest age group, who often performed at chance levels, the coefficient of reproducibility improves to 87.5% (63 of 72). Moreover, the coefficient of reproducibility for only the adult sample is an impressive 91.7% (22 of

![Boxplot of the mean scores on the composite Irony variable.](image)

**Table 1**

<table>
<thead>
<tr>
<th>Predicted patterns</th>
<th>Components</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meaning</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>Belief</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Intention</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>+</td>
<td></td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>Attitude</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>–</td>
<td>+</td>
<td>4</td>
<td>24</td>
</tr>
</tbody>
</table>

**Mean frequency of occurrence**

<table>
<thead>
<tr>
<th>Age</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Other</th>
<th>Total</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>6</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>7-year-olds</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>9-year-olds</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>9</td>
<td>15</td>
<td>30</td>
<td>17</td>
<td>15</td>
<td>96</td>
<td></td>
</tr>
</tbody>
</table>

Note. The mean frequencies were calculated for age groups based on the means of each individual’s score across all irony scenarios administered; decimal points are not included for clarity purposes.
Children of all ages performed very well on the advanced ToM measures (i.e., second-order false belief task, “strange stories” task, and “faux pas” task). Even the group of 5-year-olds answered correctly the questions examining the first-order beliefs (98%), second-order beliefs (79%), and second-order intentions (81.5%) (although this group’s mean chronological age was 5.8 years, its mean age-equivalent score on the PPVT – III was 7.1). Therefore, subsequent correlational analyses and regressions use only justification scores obtained from these advanced ToM measures, which are the only scores reflecting enough variance (see Table 3 for the frequency and percentage of justification scores).

A single score for the advanced ToM measures was needed for subsequent regression analyses. The correlations of the justification scores on the advanced ToM tasks were high: children’s justifications for the “strange stories” task correlated with those for the “faux pas” task, \( r(70) = .43, p < .001 \); the justifications for the second-order false belief task correlated with those for the “strange stories” task, \( r(70) = .38, p < .01 \), and with those for the “faux pas” task, \( r(70) = .30, p < .05 \). Considering that the different advanced ToM tasks target slightly different aspects of social reasoning, \( \alpha = .69 \) revealed by the reliability analysis provided an acceptable level for creating a single composite advanced ToM (AdvToM) score, generated by summing the justification scores across all three types of advanced ToM tasks (range 0 – 12).

In order to understand the contribution of different factors to success on the irony task, correlational analyses were performed using children’s scores from different measures. Because the adult participants were not tested on these measures, the correlations and subsequent regression analyses include only the 72 school-age participants. In the correlations summarized in Table 4, age was measured in months; raw
scores were used for vocabulary, digit-span, and prosody tasks; composite AdvToM was used for the advanced ToM measures; and composite Irony scores were used for the irony task. Table 4 illustrates that the correlations among Irony scores, AdvToM, age, and vocabulary were strong, as were those with the forward digit-span and prosody scores. However, once age, vocabulary, and digit-span scores were controlled for, the advanced ToM scores remained correlated to Irony scores only marginally, $pr = .22, p < .10$. Contrary to expectation, the correlation of the Irony scores and prosody also attenuated once age, vocabulary, and digit-span scores were partialed out, $pr = .21, p < .10$.

Table 3

<table>
<thead>
<tr>
<th>Justification score</th>
<th>Age group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5-year-olds</td>
</tr>
<tr>
<td>Second-order false belief task</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>6 (29)</td>
</tr>
<tr>
<td>1</td>
<td>11 (44)</td>
</tr>
<tr>
<td>2</td>
<td>8 (31)</td>
</tr>
<tr>
<td>Strange stories</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>9 (36)</td>
</tr>
<tr>
<td>1</td>
<td>9 (36)</td>
</tr>
<tr>
<td>2</td>
<td>7 (29)</td>
</tr>
<tr>
<td>Faux pas task</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>7 (29)</td>
</tr>
<tr>
<td>1</td>
<td>7 (29)</td>
</tr>
<tr>
<td>2</td>
<td>11 (44)</td>
</tr>
</tbody>
</table>

Factors Predicting Children’s Performance on the Irony Task

Two sets of multiple linear regressions were performed with the composite Irony scores as the dependent variable. Age (in months) and raw scores on vocabulary, digit-span, aggregate AdvToM, and prosody scores were independent variables. The first set of regressions considered the contribution of language and advanced ToM to irony understanding; the second set of regressions examined the contribution of advanced ToM and an attunement to prosody to irony understanding.

Contribution of language and advanced ToM to irony understanding. In the first set of regressions, age, digit-span, and prosody scores were entered first to ensure that the contribution of vocabulary and advanced ToM was independent of age, short-term memory, and detection of emotions from prosody. These factors were entered as a block in a single step. Subsequently, two regression analyses were performed.

First, the PPVT – III score was entered in Step 2 and AdvToM score was entered in Step 3. The first set of predictors—age, digit-span, and prosody scores—accounted for a significant amount of the variability on the Irony scores, $R^2 = .31, F(4, 67) = 7.64, p < .001$. The contribution of vocabulary over and above age, memory, and the detection of prosody was significant, $R^2$ change = .07, $F(1, 66) = 7.33, p = .009$. The contribution of advanced ToM over and above age, memory, detection of prosody, and vocabulary was modest but significant, $R^2$ change = .04, $F(1, 65) = 3.94, p = .051$. Thus, both vocabulary and advanced ToM are reliable predictors of irony understanding over and above age, memory, and the detection of prosody when considered separately.

In an alternative regression, the order of Steps 2 and 3 in the previous regression was reversed in order to investigate whether the contribution of AdvToM and vocabulary holds when these are ordered differently. AdvToM as the second predictor accounted for significant variability in Irony scores over and above age, memory, and prosody score, $R^2$ change = .07, $F(1, 66) = 7.39, p = .008$. Vocabulary, entered as the third predictor, accounted for an additional variability on the Irony scores, $R^2$ change = .04, $F(1, 65) = 3.88, p = .053$.

Results of this alternative regression confirmed those of the previous one in that advanced ToM and vocabulary are reliable and independent predictors of irony understanding over and above age, short-term memory, and the ability to detect emotions from prosodic cues.

Contribution of advanced ToM and prosody to irony understanding. Age, vocabulary, and digit span were

<table>
<thead>
<tr>
<th>Task</th>
<th>Age group</th>
<th>Forward</th>
<th>Backward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irony</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPVT – III</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward DS</td>
<td>.51***</td>
<td>.35**</td>
<td>.37**</td>
</tr>
<tr>
<td>Backward DS</td>
<td>.58***</td>
<td>.31*</td>
<td>.41***</td>
</tr>
<tr>
<td>PPVT – III</td>
<td>.62***</td>
<td>.51***</td>
<td>.45***</td>
</tr>
<tr>
<td>PPVT – III</td>
<td>.79***</td>
<td>.46***</td>
<td>.33**</td>
</tr>
</tbody>
</table>

Note. $N = 72$. AdvToM = advanced theory-of-mind Score; PPVT – III = Peabody Picture Vocabulary Test III; Forward DS = forward digit-span score of the Children’s Memory Scale; Backward DS = backward digit-span score of the Children’s Memory Scale. $* p < .05$. $** p < .01$. $*** p < .001$. 

Table 4

Correlation Matrix for Irony Score and All Other Measures
entered first in the second set of regressions to ensure that the contribution made by the remaining factors was independent of these. The three factors were entered as a block in a single step. To determine the contribution of advanced ToM and prosody over and above age, vocabulary, and short-term memory, two regression analyses were performed.

In the first regression, the AdvToM score was entered in Step 2 and the prosody score was entered in Step 3. The first set of predictors—age, vocabulary, and digit span—accounted for a significant amount of the variability on the Irony score, $R^2 = .35$, $F(4, 67) = 8.96, p < .001$. The contribution of advanced ToM over and above age, vocabulary, and memory was modest but significant, $R^2$ change $= .04$, $F(1, 66) = 3.99, p = .050$. The contribution of prosody cues identification over and above age, vocabulary, memory, and advanced ToM failed to reach significance, $R^2$ change $= .03$, $F(1, 65) = 3.30, p = .074$. Thus, when considered separately, advanced ToM is, unlike children’s attunement to prosody, a relatively good predictor of irony understanding over and above age, vocabulary, and memory.

In an alternative regression, Steps 2 and 3 from the previous regression were reversed to determine whether prosody contributes to irony understanding before AdvToM score is partialed out. Prosody as the second predictor did not account for a significant variability in Irony scores over and above age, vocabulary, and memory, $R^2$ change $= .03$, $F(1, 66) = 2.66, p = .11$. However, AdvToM score as the third predictor accounted for a significant variability in the Irony scores over and above age, vocabulary, memory, and prosody, $R^2$ change $= .04$, $F(1, 65) = 4.62, p = .035$.

The results of this set of regressions confirmed those from the previous set in that advanced ToM is a reliable predictor of Irony scores when controlling for age, vocabulary, and memory. However, the contribution of prosody to children’s performance on the Irony task failed to reach significance even when the age groups were considered separately in a series of regression analyses not reported here. Thus, the hypothesized importance of children’s attunement to prosody was not substantiated; the attunement to expressive prosody by itself is not a significant predictor of children’s overall understanding of an ironist’s mind.

Discussion

The results of the present study show a clear developmental pattern in the ability to interpret ironic utterances. The findings confirm that reflective interpretation of indirect speech, such as discourse irony, is difficult for children. Even the 9-year-old participants in this study did not reach adult skill levels, a finding that validates the predicted progression in understanding and interpretation of social-cognitive components of others’ minds beyond the preschool years. The proposed scale of social reasoning about irony incorporates a stepwise developmental sequence of gradually acquired skills. Specifically, the ability to judge the literalness of a statement precedes the ability to interpret the speaker’s mind. The present study replicates Ackerman’s (1983) and Hancock et al.’s (2000) findings that many young children who successfully reject the literal meaning of a statement have difficulty understanding the speaker’s intent. Once children can evaluate the statement, they begin to reflect on the speaker’s mental state. Our findings indicate that children become able to represent the speaker’s beliefs before they can discern his or her communicative intention. Only after correctly judging the speaker’s intention are children capable of reflecting on the speaker’s attitude. These findings lend support to Winner and Leekam’s (1991) contention that identification of the attitude behind the speaker’s choice of an ironic statement represents a higher order understanding of mind.

Sokol and Chandler (2003) proposed that the extent to which children consider active mental agency separates their early, static conception of desires and beliefs from their subsequent understanding of intentions as mental events. Ironic messages indisputably require active involvement of speaker and listener in the act of interpretation. The point of irony is to ensure that the utterance is recognized as untrue. However, the hearer’s recognition of the speaker’s intention to convey a nonbelief of the opinion expressed does not provide insight into the speaker’s choice of an untrue statement over a literal direct assertion. Recognizing the pragmatic force of the speaker’s attitude is an additional requirement for understanding irony, which continues to develop beyond the middle school years. Indeed, in the present study even adult performance was not at ceiling.

The adults were superior to children in constructing an appropriate interpretation of the speaker’s communicative intention and attitude. The findings demonstrate that children’s interpretation of these aspects of the speaker’s mind is imperfect, to say the least. No 5-year-olds were at Pattern 5 and, averaged across the eight scenarios, only two of twenty-four 7-year-olds were. Indeed, children have a long way to go before they can construct interpretations of the
speaker’s complex mental states in the way that adults do. A judgment of the speaker’s intentions and attitudes is interpretative (Sokol & Chandler, 2003), and the difficulty children experience in identifying the speakers’ communicative intentions and attitudes may stem from conceptual complexities of the representation of these mental states in addition to the structural complexities involved in their recursive application (e.g., intending someone to disbelieve the truth value of what one says or intending someone to recognize that one is just mocking something or being humorous). An increased complexity, both conceptual and structural, involves increased requirements for deciphering the sophisticated representations behind discourse irony and is likely the reason for children’s lack of a full appreciation of irony.

The findings of the present study are timely especially because of newly emerging evidence suggesting that social-reasoning skills may represent a special category of human cognition characterized by increasing sophistication, even in the face of cognitive decline. Happé, Winner, and Brownell (1998) demonstrated that even though performance on tasks requiring nonmental reasoning (i.e., stories with physical content and jumbled texts) declines in later adulthood, performance on tasks requiring mental-state reasoning (i.e., ToM) remains undiminished and may even improve with age. Along with the identification of advanced aspects of social reasoning documented thus far, our study contributes importantly to the identification of mechanisms involved in reasoning about dyadic communication.

Linguistic ability is one of the major factors in children’s developing skills of interpreting other minds in discourse. It has been shown to facilitate acquisition of theory of mind in early childhood (Austingon & Filippova, 2005; Milligan et al., 2007). The findings of this study reveal that language competence plays an important role even in school-age children’s detection of advanced aspects of others’ minds. The regression analyses show that both language competence, measured by the receptive vocabulary test, and advanced ToM abilities, measured by justifications from advanced ToM tasks, contributed independently to children’s understanding of the ironic speaker’s mind. Given that the scope of both language and advanced ToM measures was limited (i.e., receptive vocabulary and coded justifications, respectively), the results provide adequate support for the hypothesis that language and ToM make independent contributions to children’s interpretation of irony. Language skills contribute to irony understanding in an important way, even after controlling for age, attunement to prosody, and advanced ToM skills. Our findings demonstrate that ToM and language make unique contributions to children’s interpretation of irony over and above the impact of age, memory, and attunement to expressive prosody.

However, there was no evidence that attunement to paralinguistic cues in prosody made a unique contribution to children’s overall ability to interpret an ironist’s mind once age, memory, and advanced ToM skills were taken into account. This is not to suggest that prosody plays no role in children’s understanding of irony. To our knowledge, our study was the first to use an independent measure of prosody to assess the contribution of such skills to children’s success at deciphering ironic messages, rather than manipulating acoustic features within the irony task (e.g., Capelli et al., 1990; de Groot et al., 1995; Dews et al., 1996). Thus, reconciling our nonfinding with a body of conflicting findings on children’s reliance on intonation in deciphering verbal irony would not be warranted. We would argue, however, that although children can identify a speaker’s emotions from nonverbal intonational cues when semantic cues are absent, they are less likely to use these cues when semantic cues are available. Morton and Trehub (2001) demonstrated that young children judge a speaker’s feelings from message content rather than intonation when these sources provide conflicting cues to emotion. Thus, children’s bias toward message content may have led them to discount the relevance of ironic prosody, even if they might have been able to match the emotions with the acoustic features in the independent prosody task we used in our study.

In conclusion, verbal irony represents a valid means for studying children’s developing understanding of advanced aspects of people’s minds in social context. An important contribution of this study is the finding that children’s language skills continue to contribute to their abilities to interpret other minds even beyond the preschool years.

References


**Appendix A: Sample of an Irony Story**

Robert is the new player on his school’s soccer team. He is really excited about being on the team. Robert’s best friend Oliver also plays on the team. During his first game, Robert misses the chance to score several easy goals. His team loses the game.

After the game, Oliver says to Robert: “You sure ARE a GREAT scorer!”

**Comprehension Questions:**
1. Did Robert help his team to win the game? [correct answer: no]
2. What did Oliver say to Robert? [exact retell, a paraphrase, or the correct gist]

**Meaning:**
3.1. Does Oliver mean that? [correct answer: no]
3.2. What does he mean [acknowledged meaning opposite to the literal statement]

**Belief:**
4. Does Oliver think Robert is a great scorer? [correct answer: no]

**Communicative intention:**
Oliver said to Robert: “You sure ARE a GREAT scorer!” [correct answer: no]

**Motivation/Attitude:**
6. Why did he say that? [score 4 being funny, teasing, joking
3 blame, criticize, make him feel bad
2 wanted him not to do it again
1 because he was bad, disappointed
0 he said it by accident; ‘I don’t know’]
(see also Appendix C for a detailed coding scheme)

**Appendix B: Samples of Advanced Theory-of-Mind Tasks**

**Second-Order False Belief Task**

*Tom’s crayon (Astington et al., 2002)*

This is a story about Tom and his sister Mary. This is Tom and this is Mary. One day, Mary and Tom are in Tom’s room. Tom is drawing with his favorite crayon. “I’m going to get a drink of water,” says Tom. He puts his crayon into the desk and leaves the room. Tom knows that his sister likes to play tricks on him, and so he peeks back around the door at Mary. When Mary sees that Tom has left, she decides to play a trick. She takes Tom’s crayon from the desk and puts it in the basket. Tom sees her do this but Mary can’t see Tom!

1. Can Tom see Mary?
2. Where does Tom think the crayon is?
3. Does Mary think that Tom can see her?
4. Where does Mary think Tom will look for the crayon when he comes back into the room?
5. Why does she think this?

**Strange Stories Task**

*Simon and Jim (Happeé, 1994)*

Simon is a big liar. Simon’s brother Jim knows this. He knows that Simon never tells the truth!

One day Simon stole Jim’s baseball bat. Jim is sure that Simon has hidden it somewhere. He can’t find it. He’s really mad. He finds Simon in his room and he
says: “Where is my baseball bat? You must have hidden it either in your closet or under your bed. I’ve looked everywhere else. Where is it, in the closet or under the bed?” Simon tells him the bat is under his bed.

Explanation: Why will Jim look in the closet?

Faux Pas Task

Rocket (Banerjee, 2000)

This is Sally, and this is her friend Nick. Nick has painted a picture of a rocket for a special exhibition in the school. Nick is really proud of his picture and thinks it’s great. Anyway, they’re walking home from school when they bump into Sally’s friend, Amy. Amy has never met Nick before. ‘Hi,’ says Amy. Sally asks her, ‘Hi Amy, have you been to the exhibition in the school?’

‘Yes,’ says Amy, ‘it’s quite nice, except there’s that really horrible painting of a rocket. I think it’s a really bad painting!’

Control: Did someone say something they shouldn’t have? Who was it? What did Amy say?

Identification of a faux pas: Why shouldn’t Amy have said that?

First-order belief: Did Amy know it was Nick who painted the picture of the rocket?

Second-order belief: What does Nick think that Amy thinks about his painting?

Second-order intention: Did Amy mean to upset Nick?

Justification: How do you know [child’s response]?

Appendix C: Coding Scheme for the Motivation/Attitude Question

<table>
<thead>
<tr>
<th>Score</th>
<th>Type of justification</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Inappropriate justifications</td>
<td>Answers irrelevant to the identification of the speaker’s motivation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S/he said it by accident; she got mixed up with words</td>
</tr>
<tr>
<td>1</td>
<td>Simple surface-level justifications</td>
<td>Reference to story facts that have relevance to the identification of the speaker’s motivation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Because she played with the boxes; because she wetted her; because he broke the pitcher/plate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S/he was mad, upset, disappointed, happy, glad, surprised, etc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S/he was being kind, mean, etc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Because that’s not what you do in a store but on a playground</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S/he was not supposed to do that; because it wasn’t nice/good</td>
</tr>
<tr>
<td>2</td>
<td>Identification of the speaker’s first-order intention</td>
<td>Reference to information that has an implication for another person’s behavior, rather than mental state</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S/he wanted him not to do it again; so she doesn’t do it the next time</td>
</tr>
<tr>
<td>3</td>
<td>Identification of the speaker’s second-order intention</td>
<td>Identification of the speaker’s intention to affect another person’s mental state</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Get the message across; make aware; acknowledge; blame; criticize; praise; compliment; thank make him/her feel good/better/bad about what s/he did</td>
</tr>
<tr>
<td>4</td>
<td>Identification of an attitude</td>
<td>Reference to the speaker’s attitude toward the situation or the pragmatic function of irony that goes beyond the identification of the speaker’s intention</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Being funny; joking; enlightening the situation; adding humor contrasting; mocking; exaggerating; emphasizing; being sarcastic; being ironic</td>
</tr>
</tbody>
</table>