2013-05-01

Speaker Information as a Cue to Irony Perception in Middle Childhood

Whalen, Juanita Marie

UNIVERSITY OF CALGARY

Speaker Information as a Cue to Irony Perception in Middle Childhood

by

Juanita Marie Whalen

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

CALGARY, ALBERTA

APRIL, 2013

Abstract

In this thesis I examined the process of irony appreciation across middle childhood with three goals: 1) to identify potential speaker cues that are useful for irony interpretation, 2) to ascertain developmental trends in processing that relate to the integration of these speaker cues, and 3) to determine whether children exhibited evidence of interactive processing. Two experiments were conducted in which 5- to- 6-year-old children, 7- to-8-year-old children, and adults viewed puppet shows in which a speaker cue was manipulated. The puppet shows ended in a literal or ironic statement and measurements of participants' behavioral and processing responses were recorded.

In Experiment 1 the speaker cue manipulated was sibling relationship status; the speaker and target in the story were either siblings or had just met. Adults and children both showed a trend in their response latencies to respond more quickly when a sibling made an ironic criticism than when a stranger did. However, this finding was not robust. In Experiment 2 the speaker cue manipulated was the speaker's perceived similarity to the participant. The participant identified a speaker that was similar to them, and one who was dissimilar to them; each of these two speakers were used in half of the stories the participants heard. The results of this experiment showed that adults benefited from the presence of a similar speaker in some aspects of irony interpretation, but children did not. Across both experiments, evidence from eye gaze patterns supported interpretations that are consistent with interactive processing accounts of irony. Implications for the use of speaker cues in irony understanding and the interpretation of results within modular versus interactive processing accounts are discussed.

Acknowledgements

I extend my deepest and most sincere thanks to my supervisor, Dr. Penny Pexman. Her thoughtful and supportive guidance have benefited me more than any other aspect of my graduate training, by far. I thank her for her incredible mentorship and support. I feel immensely lucky to have found Penny, and luckier still that she tolerated me for so long.

I also thank my committee members, Dr. Susan Graham and Dr. Susan Boon, for their time and attention, and for their helpful feedback at all stages of this thesis. I also thank the external members of my examination committee, Dr. Adrienne Kertzer and Dr. Jeffrey Hancock.

I thank the adult participants, and the child participants and their families, who participated in these experiments. I also thank the researchers who established and maintain the C.h.I.L.D. database, from which the child participants were drawn. This database is a remarkable resource for those of us conducting developmental research and its existence in the Psychology Department makes participant recruitment immeasurably more feasible.

This research was supported by a SSHRC doctoral fellowship, and I am very grateful to have had the financial support to be able to pursue this degree and to be able to conduct this research.

Thank you to the Pexmanites, my labmates (past and present), for the support throughout this degree. For being a sounding board, for assisting with all my little methodology and research queries, and for pretending to care about puppet research. In particular, I am indebted to Gemma Leonard for assistance with participant recruitment

and booking, data collection, video coding, data coding and data entry. I also extend my sincere thanks to Ian Hargreaves for his ongoing support with statistical and technical queries. I still owe him. I thank Kristy Kowatch for her assistance with video coding, and thank Catie Phillips for her thoughtful comments on talks, posters and on this document.

Thanks to the wonderful ladies who jumped in to provide childcare when I had participants to squeeze in or had other deadlines to meet: Dr. Melissa Boyce, Dr. Laurie Ching, and Lisa McCarthy. Also, thanks to my sisters, my extended family, and my friends for their continued support; your unfailing support is simultaneously touching and a little bit crazy.

Finally, the best way I can offer thanks to my family is to stop abandoning them on the weekends, but allow me to also thank them here. Clayton, thank you for stepping up and doing the myriad things I had to drop to finish this document: thanks for taking time off with our sick kid, for doing the shopping, the cleaning, the cooking, the schlepping to and fro, the wrangling, the pep-talking, and for just generally tolerating me while I placed you at the bottom of my priority list. Lola, thank you for indulging my absenteeism, for the cuddles, movie breaks and dance parties, for forcing me to learn how to juggle things that are hard to juggle, for introducing me to Angus, and for your sweet renditions from *Revolver*.

Dedication

To Clayton and Lola, for making the bad days good and the good days better.

Table of Contents

List of Tables

List of Figures

General Introduction

Imagine you are at a local park having a picnic with your family. You can hear children playing on the playground behind you, and although you cannot see them you hear one child say to another "Nice going, Marcus!" Your son looks at you and asks "Why did that girl say that to him?" Since you didn't see what prompted the utterance, and know nothing about the children involved, it might be difficult to give your son a definitive answer about what the speaker's intent was when she was addressing her playmate, Marcus. It is possible that the girl wanted to encourage or compliment Marcus's efforts on something and was being literal with her remark. Alternatively, it is possible the girl felt Marcus failed to do something well and was criticizing his performance indirectly, therefore speaking figuratively. Additional information might help you determine whether the girl was being literal or figurative in her utterance: whether the children were younger or older, whether they knew one another or had just met at the playground, and whether they had any expectations in advance about Marcus' performance. Certainly, the tone of voice in which the utterance was made should be helpful. Most importantly, in order to feel certain about the speaker's intention in this case you would benefit from having seen the event unfold. Determining the speaker's intent with this utterance is difficult enough for you, but for a child like your son it is additionally taxing because children lack the experience to understand the importance of the multiple social, linguistic and contextual cues present, and may have limited cognitive skills for coordinating that information in real time. The present study will focus on how situational cues shape children's processing and understanding of ironic language.

Verbal irony is a form of figurative language in which the speaker intends to convey a meaning that differs from the spoken message (e.g., Gibbs & Colston, 2012). Many sources characterize irony as a counterfactual expression, where the intended meaning is the opposite of the literal meaning of the utterance (e.g., Gibbs & Colston, 2012; Roberts & Kreuz, 1994). Much of the literature, particularly in developmental psychology, examines the use of sarcasm specifically, which is not only counterfactual, but is also critical and is directed towards a particular person or target (Lee & Katz, 1998). There are several aspects of verbal irony that have been examined and there are many more than remain to be examined. In this thesis I address three different components of irony appreciation. First, I examine the cues relevant to irony appreciation. Context is deemed to be critically important in irony understanding (Clark, 1992), but context can vary greatly, with context being very rich in some instances and very impoverished in others. Regardless of how impoverished a contextual setting is, there is always a speaker involved when an ironic utterance is being made. As such, this research examines cues relating to the speaker. Second, I examine how these cues might be integrated during the processing of the remark. Much of the developmental irony literature has focused on the achievement of comprehension; that is, measuring the end product of interpretation. A greater understanding of how processing unfolds is required to assist with theory refinement. Third, I examine relationships between interpretation, processing and other developmental achievements that may relate to irony appreciation, such as theory of mind. The subsequent sections outline past literature that is relevant to each of these three components. First, however, I will review the goals of ironic discourse and past findings about children's understanding of ironic language.

In present day communication adults use the various forms of irony in about 8% of conversational turns (Gibbs, 2000) to achieve many different functions. For example, irony can be used to establish rapport or solidarity (Colston, 1997), to tease (Kreuz, Long, & Church 1991; Pexman & Zvaigzne, 2004), to add humour (Colston & Keller, 1998; Colston & O'Brien, 2000; Dews & Winner, 1995; Dress, Kreuz, Link, & Caucci, 2008; Jorgensen, 1996; Kreuz et al., 1991; Kumon-Nakamura, Glucksberg, & Brown, 1995; Pexman & Olineck, 2002b, Roberts & Kreuz, 1994), to protect the target's feelings by being polite (Brown & Levinson, 1987), to mute criticism (Dews & Winner, 1995; Pexman & Olineck, 2002a), and to express negative attitude or criticism (Colston, 1997; Roberts & Kreuz, 1994).

Evidence from a laboratory setting suggests that children begin to use counterfactual irony as early as age 7 (Whalen & Pexman, 2010), and one naturalistic study has provided evidence of 4- to 6-year-olds using forms of irony in a family setting (Recchia, Howe, Ross, & Alexander, 2010). By age 5 or 6 most typically developing children demonstrate a basic understanding of irony (Ackerman, 1983; de Groot, Kaplan, Rosenblatt, Dews, & Winner, 1995; Dews et al., 1996; Filippova & Astington, 2008; Hancock, Dunham, & Purdy, 2000; Harris & Pexman, 2003; Winner & Leekam, 1991). However, a more thorough appreciation of irony continues to unfold across middle childhood and into adolescence and adulthood (Capelli, Nakagawa, & Madden, 1990; Demorest, Silberstein, Gardner, & Winner, 1983; Filippova & Astington, 2008). The understanding of other forms of irony develops later in childhood: ironic compliments (i.e., the delivery of a complimentary message via negatively-worded language) are not usually understood as early as ironic criticisms (i.e., the delivery of a critical message via

a positively worded utterance; Hancock et al., 2000; Harris & Pexman, 2003; Pexman & Glenwright, 2007; Pexman, Glenwright, Krol, & James, 2005). In addition, children do not begin to grasp the intended humour in ironic criticisms until age 9 or 10 (Dews et. al., 1996; Harris & Pexman, 2003; Pexman et al., 2005).

Several layers of cognitive mastery must be reached in order for a fuller appreciation of this speech form to take hold (Miller, 2012). First of all, the listener must identify that the speaker's statement is false, and that it is not merely a mistake but is in fact intended to be false. Second, the listener needs to understand that the false utterance is not meant to be an act of deceit, such as a white lie intended to protect someone's feelings, but instead is meant to communicate an attitude about the topic at hand. Next, the listener needs to understand the speaker's pragmatic goal in making the utterance (Filippova & Astington, 2008; Miller, 2012). As noted above, these pragmatic goals may include the desires to be critical, humorous, or polite, just to name a few of the common discourse goals.

Cues to Ironic Intent

Part of the challenge children face in mastering irony appreciation arises due to the fact that a wide assortment of cues might be relevant in arriving at an ironic interpretation of an utterance. Listeners gather information from a variety of sources when deciphering the intent of a remark. Contextual information is useful in assisting a listener in determining the appropriate interpretation when a formally ambiguous utterance is made (Clark, 1992), and it is well established that a strong incongruity between the context and the statement can facilitate irony perception (Colston, 2002; Colston & O'Brien, 2000; Ivanko & Pexman, 2003; Pexman & Olineck, 2002a; Pexman,

Whalen, & Green, 2010). Remarks that make reference to social norms or to unmet expectations also increase the salience of the ironic interpretation because they highlight discrepancies that exist between those expectations and the reality (Creusere, 2000; Gibbs, 1986; Kreuz & Glucksberg, 1989; Kumon-Nakamura et al., 1995). Further, ironic tone of voice, or changes in intonation, can signal ironic intent (Bryant, 2011; Bryant & Fox Tree, 2005; Capelli et al., 1990; Winner, 1988). Different combinations of these cues can be relevant in different settings, and it seems unlikely that any particular cue is always a marker for ironic intent. Instead, listeners coordinate information from several sources when contemplating speaker intent.

Previous research suggests that both children and adults consider information about the speaker to be relevant to an ironic interpretation of that speaker's remark. Adults use information from stereotypes concerning a speaker's occupation to infer ironic intent (Katz & Pexman, 1997). For instance, irony is detected more readily in scenarios when the speaker is a comedian than when the speaker is a priest. Adults also use information about the speaker-target relationship to infer ironic intent, detecting irony more readily in close speaker-target relationships (Pexman et al., 2010; Pexman & Zvaigzne, 2004). Research has established that children can use a speaker's intonation and facial expression to assist in understanding irony (Winner, 1988) and can use speaker personality trait information to understand ironic intent (Pexman, Glenwright, Hala, Ivanko, & Jungen, 2006). That is, children detect ironic criticism more readily when the speaker is described as a mean person. Knowledge about the speaker can provide insight into the speaker's attitude, and in turn, knowledge about the speaker's attitude can help the listener determine ironic intent. Even the relative expertise of the speaker can impact

children's understanding of ironic language. When participants heard stories in which mothers or siblings made ironic remarks towards a child, the participants grasped ironic intent more readily when the relationship between the characters was a mother-child relationship rather than a sibling relationship (Massaro, Valle, & Marchetti, 2012). The authors suggested that adult speakers have conversational expertise that allows them to scaffold ironic language learning, and listeners who interpret these scenarios are sensitive to that conversational expertise when judging whether the speaker was being literal or ironic (Massaro et al., 2012).

Each of the above cues---intonation, context incongruity, and speaker trait information---has been demonstrated to assist with irony comprehension in some way. However, theoretical accounts to explain the coordination of relevant cues have been incomprehensive. There are several theories that adequately explain some particular element of ironic language understanding, but few that address development or the process of meaning resolution that unfolds in irony interpretation.

Processing Accounts of Irony Appreciation

The process of irony appreciation is believed by some theorists to proceed in a stage-like fashion. This modular approach has roots in Grice's (1975) standard pragmatic model whereby an ironic interpretation is only considered after the literal interpretation has been examined and rejected as a viable interpretation based on the surrounding context. Similar to the Gricean view, Giora's (1997) graded salience hypothesis posits the literal meaning is the most salient message meaning, and is rejected in favor of an ironic interpretation only after the literal interpretation is found to be insufficient. This two-stage process to arrive at an ironic interpretation of an utterance leads to the expectation

of longer processing times for irony than for literal language.

In contrast to these literal-first processing accounts are interactive processing approaches. Gibbs' (1986) direct access view suggests that ironic interpretations of an utterance can be considered from the earliest moments of processing. The ironic interpretation can be accessed just as readily as the literal one if context suggests an ironic interpretation. Another interactive processing view is the parallel constraint satisfaction approach (Katz, 2005; Pexman, 2008; Thagard, 2000). This account suggests that listeners make use of multiple cues in their environment, and consider these cues in parallel, in order to comprehend ironic language. During the process of constraint satisfaction the user considers several elements that are accepted or rejected in relation to the problem space (Thagard, 2000). Elements that cohere have positive constraint between them and elements that do not cohere have negative constraint. This approach has been applied in many topic areas. In the case of discourse comprehension, the problem needing resolution may be an ambiguous utterance, and the elements under examination may be cues to interpretation. As each cue is considered it is accepted as plausible or rejected as implausible. Importantly, this process does not happen in a serial fashion, but instead is carried out in parallel. The combination of cues that leads to the clearest solution (i.e., understanding of the utterance) constitutes constraint satisfaction. The available cues are considered rapidly, and in parallel, until the perceiver arrives at a solution that is the best fit for the cues considered (Pexman, 2008).

The literal-first processing accounts have garnered some support from reading studies, with some studies reporting longer reading times for ironic statements than for literal statements (e.g., Dews & Winner, 1999). Other reading studies have provided

support for interactive accounts, by showing that ironic statements do not necessarily take longer to process than literal statements (e.g., Gibbs, 1986). Priming studies have also been used to explore the potential processing differences in ironic and literal language (e.g., Akimoto, Miyazawa, & Muramoto, 2012; Giora & Fein, 1999). These studies have all involved written language and the focus has been on adult samples. As such, we know very little about how children process ironic language and about processing of verbal irony in face-to-face settings. Measuring processing during spoken language is slightly more challenging than doing so with text. However, eye movements have been used reliably as an indicator of auditory language processing since the mid 1970's (Tanenhaus & Trueswell, 2005); it is believed that eye gaze is directed towards the elements of the environment that are currently being considered in interpretation (Trueswell & Gleitman, 2004). As such, eye gaze is believed to be a reliable indicator of processing, even outside of text-based settings. In one of the only studies to examine processing of ironic language in children, Climie and Pexman (2008) demonstrated that children access cues to irony, and consider the ironic meaning, early in the comprehension process. Using an eye gaze procedure Climie and Pexman demonstrated that children as young as age 5 do not first consider the literal interpretation of an ironic utterance, but instead contemplate the ironic interpretation from the earliest moments in processing. This evidence suggests that when entertaining the ironic interpretation early in processing, children may be incorporating information about the context and utterance together, as suggested by interactive processing accounts like the parallel constraint satisfaction approach. In order to more clearly define the parameters of the constraint satisfaction theory as a description of children's irony processing, researchers need to identify exactly which cues are relevant

for children for perception and appreciation of irony, and to understand when and how children use those cues in interpretation. The present study aimed to do this, with the ultimate goal of theory refinement.

Individual Differences Relating to Irony Appreciation

When an individual understands the mental states of others, they are said to have a theory of mind. There are several milestones in theory of mind mastery. One of these milestones includes merely knowing that another person holds particular beliefs about physical objects or actions in the world that can contradict one's own beliefs about those things; this constitutes a first-order theory of mind, which is normally mastered between the ages of 4 and 5 (Wellman, Cross, & Watson, 2001). A subsequent milestone involves also knowing that another individual can have beliefs that are based on their knowledge of your beliefs or the beliefs of a third party. This more advanced achievement relies upon one's understanding of the mental states of others rather than an understanding of actual physical event or objects; this ability constitutes a second-order theory of mind and it is not achieved until a later age. The precise age of second-order mastery varies depending on context, but many typically-developing children reliably exhibit second-order mental state reasoning starting at age 6 (e.g., Astington, Pelletier, & Homer, 2002). Since ironic language appreciation involves making an inference about an underlying belief that is not explicitly stated, it seems plausible that there would be a link between irony comprehension and theory of mind. In fact, several recent studies have explored the potential relationship between theory of mind and irony.

Filippova and Astington (2008) found that advanced theory of mind mastery impacted children's understanding of irony. The authors concluded that this advanced

mental state reasoning ability was essential to understanding the ironist's attitude about a situation (Filippova & Astington, 2008). Similarly, Massaro et al. (2012) found second-order false belief to be correlated with irony understanding among 6- to 10-year-olds, both in conversations between mothers and children, and in conversations among siblings. It should be noted that a theory of mind might not be necessary for all aspects of irony understanding. It is possible, for instance, to understand that a speaker has stated something false or counterfactual without having any insight into the speaker's mind. However, theory of mind is thought to be relevant to understanding the *intention* of the ironist, that the speaker's counterfactual statement was meant to be interpreted as irony and not interpreted as a lie or a mistake (Akimoto et al., 2012).

The Present Study

In the present study I conducted two experiments to explore the effectiveness of different types of speaker information as cues to children's interpretation of ironic criticisms. In the first experiment I explored how the presence (or absence) of a sibling relationship between speaker and target influenced children's irony appreciation and processing. In the second experiment I explored how the speaker's perceived similarity to the participant impacted the participant's irony appreciation and processing. To assess irony appreciation the experiments outlined here used puppet shows in which an important speaker cue was manipulated. Overt judgments of the speakers' intent and belief were recorded, and participants' eye gaze and reaction times were also tracked to provide clues about the cognition involved in the judgment process. Also included were individual differences measures that provide insight about the cognitive, linguistic, and social skills that are related to children's irony appreciation.

Experiment 1

The purpose of this experiment was to investigate whether children can use information about sibling relationships to support inferences about speaker intent for irony. Previous research with adults has demonstrated that close speaker-target relationships facilitate perception of the pragmatic functions of verbal irony (Pexman & Zvaigzne, 2004). The shared common ground between people in a close relationship provides clues about speaker attitude (Gibbs, 1986; Pexman & Zvaigzne, 2004), which facilitates the metarepresentational inferences involved in irony comprehension. One might assume the same closeness facilitates irony perception with children. However, previous research suggests that this is not always the case. Limited research has explored these social factors, but one study has demonstrated that children do not appreciate irony more readily when used between friends than when used between strangers (Pexman et al., 2005). A possible explanation for this finding is that children hold a different belief than do adults about the role of irony in close relationships. Whereas adults tend to think that ironic teasing is appropriate in close relationships, children may disagree. Indeed, in the Pexman et al. (2005) study children's responses suggested that they felt it would be mean to speak nonliterally or indirectly to a friend, and that in order to be kind to one's friends one should speak directly. Although friendships are an obvious instance of a close relationship they are not the only type of dyadic relationship that might create an environment suitable for the use of ironic language.

Sibling relationships may offer the closeness found useful in prior adult studies, and may be particularly suited to use of ironic expression among children. Children may feel it is appropriate to tease siblings whereas it is not appropriate to do so with friends.

There are a number of critical differences between sibling relationships and peer friendships that might permit more teasing between siblings. First, siblings spend as much, or more, time together as peers (DeHart, 1999; McHale & Crouter, 1996), so there is ample opportunity to learn about language and behavior from siblings, including the opportunity to witness, and practice, irony use. Second, siblings share more history and common ground (DeHart, 1999). This common ground may ensure that children in the same household get exposure to some of the same types of experiences and language, and the greater common ground that is established among siblings could be very useful in determining the possible interpretations of ambiguous messages. Third, siblings share more intimacy than do peer dyads (DeHart, 1999). This increased intimacy builds on the notion of common ground, and may provide children with rich personal experiences with their siblings that can later be referenced by ironic remarks. Fourth, siblings can trust their relationship will not end readily, since it is not a relationship formed by choice (Cutting & Dunn, 2006). Therefore, children may not have to work as hard at adhering to politeness norms and may feel freer to be critical in the sibling relationship. Finally, a certain level of hostility can be tolerated in a sibling relationship whereas hostility would be more rare in children's friendships (Cutting & Dunn, 2006). The increased hostility and tension may allow siblings to use ironic language as a mediation tool to either enhance, or reduce, conflict. Taken together, these relationship characteristics suggest that children may experiment with risky forms of language such as verbal irony in the context of a sibling relationship.

Sibling relationships provide a rich training ground for later peer relationships. Children are exposed to important situations in their daily lives with siblings that help

them gain experience with considering the thoughts and perspectives of another person (Miller, 2012). Social cognitive skills are observed within sibling interaction earlier than within peer interaction (Dunn & Dale, 1984). These skills are essential for navigating a wide variety of situations, including the use of pragmatics in face-to-face communication. Irony appreciation relies on mastery of certain cognitive skills, such as the ability to consider more than one outcome, to shift perspective, and to consider a variety of cues in interpretation. Mastery of ironic language also relies upon social experience (Pexman & Glenwright, 2007). It is likely that each of these cognitive and social skills is fostered in relationships in early childhood. In fact, one study has demonstrated that children with siblings have more positive social evaluations than do only children (Kitzman, Cohen, & Lockwood, 2002). The earlier emergence of these social-cognitive skills in sibling relationships further suggests that ironic language might be perceived and processed more readily when the speaker and target are siblings. If children assume that ironic language is probable in a sibling relationship then this information can be used to cue speaker intent.

A comparison of rates of relational aggression between siblings and peers has shown that preschoolers use more relational aggression with siblings, but by middle childhood children use less relational aggression with siblings and more with peers (Stauffacher & DeHart, 2006). Very young children have more opportunity to use aggressive behavior with siblings, but also may deem it more appropriate to behave aggressively with siblings than with friends. This is further evidence that children might use critical forms of irony, like sarcasm, frequently with siblings from a young age. The security of the sibling relationship may permit children to speak critically or aggressively

with the use of irony without fearing the ironic expression will have dire consequences.

My expectation is that the sibling relationship type is well suited to ironic language; children are likely to tease their siblings and as such, might be more likely to use irony with siblings than with friends or with strangers. Thus, the presence of a sibling relationship in a scenario might be a cue to ironic intent. In order to test this possibility, the stories used in Experiment 1 contained information about whether the characters (puppets) involved were siblings or had just met for the first time, e.g., "Shannon and Lisa are sisters. They play on a soccer team together. It is the last few minutes of the game. Lisa kicks the ball, missing the net entirely. Shannon says *'That was a great play.'*"

Predictions

Adult participants were also included in the present study to provide a baseline for processing speed and eye gaze behavior, and to indicate whether mature speakers might also use the sibling relationship cue. Despite this inclusion, it was not prudent to make direct comparisons between adult and child participants (see Results for further description), so predictions focused on child data only.

1. Children in the older age group (7- to 8-year-olds) will demonstrate greater accuracy for ironic criticism understanding than will children in the younger age group (5- to 6-year-olds). This effect was anticipated for assessments of speaker belief, speaker intent, and humorous intent.
2. If sibling relationship information is a cue to ironic intent, then overall accuracy will be higher on each of the behavioral measures of irony appreciation for the sibling conditions than for the stranger conditions.

3. Children in the older group will have faster response latencies to ironic criticisms than will children in the younger group.
4. If sibling relationship information is a cue to ironic intent, then response latencies to ironic criticisms in the sibling conditions will be faster than those in the stranger conditions. If response latency differences between sibling and stranger trials emerge in the earlier stages of processing, then this finding would be taken as evidence for earlier consolidation of this cue and would therefore indicate that an interactive account of processing could describe processing under these conditions.
5. The accuracy of the first fixation during ironic trials will be higher for children in the older group than for children in the younger group.
6. If the literal interpretation is considered more often early in processing, as modular processing accounts would predict, then the accuracy of the first fixation to the correct object during ironic trials will be lower than that during the literal trials. However, if the ironic interpretation is considered as readily as the literal interpretation early in processing, as interactive processing accounts would predict, then the accuracy of the first fixation during ironic trials and literal trials will be equal.
7. If sibling relationship information is a cue to ironic intent, then the mean number of fixations to the correct object in the ironic trials will be higher in the stranger condition than in the sibling condition, since a greater number of looks conveys more uncertainty in decision-making.

8. If sibling information is a cue to ironic intent, then the proportion of total fixation duration dedicated to the correct object will be higher in the sibling condition than in the stranger condition.

Method

Participants

Participants were 27 five-and six-year-old children (*M* = 5.49 years, *SD* = 0.50, 14 female), 20 seven- and eight-year-old children (*M* = 7.58 years, *SD* = 0.51, 8 female), and 20 adults (*M* = 19.96 years, *SD* = 1.64, 19 female). The child participants were recruited from the community in a large Canadian city and were tested in a lab on the university campus. Recruitment was restricted to children who had at least one sibling to ensure all participants could be expected to know how siblings might typically speak to one another. The adult participants were recruited from an undergraduate research participation system and were tested in the same lab on campus. Twenty-one additional participants were tested but were not included in the final analyses: Ten participants were excluded from analyses due to experimenter error with the testing protocols, eight were excluded due to problems with the video recording, and three were excluded due to poor overall performance on the manipulation check (described further below). The child participants received a prize for participating in the study and adult participants received partial credit towards a Psychology course.

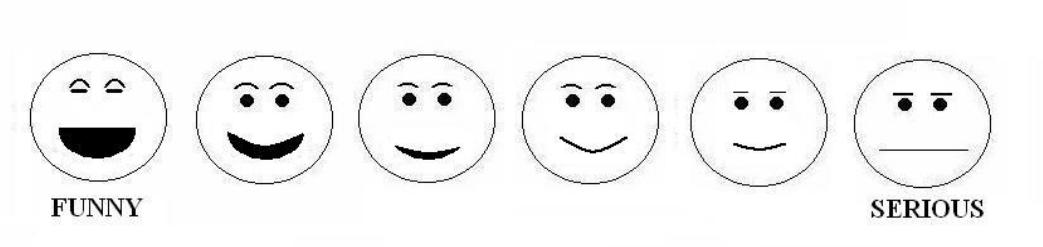
Materials

Puppet shows. Ten puppet shows, each involving two different puppets and props, were used to depict stories involving two characters: a speaker and target that were siblings or a speaker and target that were strangers. The puppet shows involved situations

that are familiar and enjoyable to most children, such as playing video games or going snowboarding. Each story was four to six sentences long. Speaker-target relationship information was provided at the start of each puppet show. For example, in the sibling condition, "Shannon and Lisa are sisters. They play on a soccer team together…", and in the stranger condition, "Shannon and Lisa just met. They just started playing for the same soccer team…" Each puppet show ended with an evaluative statement made by the speaker puppet. The evaluative statement was an ironic criticism for six of the stories, a literal criticism for two of the stories, and a literal compliment for two of the stories. The language type employed in the evaluative statement was counterbalanced across all stories and story presentation was randomized across the 10 protocol versions used. Appendix A contains sample puppet show stories, with the possible language endings.

For judgments of speaker intent two small stuffed toys and a small answer box were used. Participants made their judgments by choosing a toy and placing it in the small, open box. One toy was a soft, smiling duck, which was referred to as the "nice duck" and the other was a scowling shark with sharp teeth, which was referred to as the "mean shark." For judgments of humorous intent a scale depicting face drawings was used. The scale options are 1 = very serious, 2 = serious, 3 = a little bit serious, 4 = a little bit funny, 5 = funny, and 6 = very funny. Each response option was accompanied by a corresponding serious or funny face (Figure 1).

Figure 1. Humour Rating Scale



Language measures. Two language measures were used. The first was The Peabody Picture Vocabulary Test (PPVT-4; Fourth Edition, Form A; Dunn & Dunn, 2007). The PPVT-4 (Dunn & Dunn) was administered to all participants to measure receptive vocabulary as an indicator of linguistic proficiency. Administration of this test took a maximum of approximately 10-15 minutes per participant. The PPVT-4 is suitable for individuals aged 2 years, 6 months to 90 years, making it appropriate for use with both child and adult participants. For the purpose of this study, the raw PPVT-4 score was used.

The second language measure was the Children's Communication Checklist (CCC; Bishop, 1998). This measure was administered to the parents of child participants as a measure of pragmatic language understanding. This measure contains 9 subscales: speech output, syntax, inappropriate intonation, coherence, stereotyped conversation, use of conversational context, conversational rapport, social relationships and interests. A pragmatic composite score is obtained by combining a selection of these subscales. For the purpose of the present study, the pragmatic composite score was used.

Theory of mind measure. Two stories designed to measure children's false belief understanding were administered at the end of each session to the child participants only.

Each of these stories is approximately 13 sentences long, and was accompanied by colour drawings, in a storybook format. These stories, which were modifications of the Perner and Wimmer (1985) and Sullivan, Zaitchick, and Tager-Flusberg (1994) scenarios, were used to establish children's first- and second-order false belief understanding. Five questions were presented throughout the theory of mind stories. The first two questions were used as comprehension controls, the third question assessed first-order false-belief, the fourth assessed second-order false belief, and the fifth was used to solicit the participant's justification for the second-order false-belief. The text from the false belief stories can be found in Appendix B. Each question was scored dichotomously as being correct or incorrect, and participants were given one point for each correct question. A composite score out of 10 possible points (5 each for the two stories, one for each question asked) was used as an indicator of theory of mind ability.

Sibling questionnaire. Parents of child participants completed the brief version of the Sibling Relationship Questionnaire (SRQ; Furman & Buhrmester, 1985). This questionnaire includes 39 items that are used to obtain a measure of sibling relationship quality. The questions on this questionnaire produce loadings for four factors: warmth/closeness, relative status/power, conflict, and rivalry. Items relating to the rivalry score were less relevant to the main focus of this study, so these items were eliminated from the scale for the present study.

Demographic questionnaire. Parents were provided with a brief demographic questionnaire to elicit information about the participants' number of siblings, birth order, sibling genders, and other relevant factors (Appendix C).

Procedure

Participants met with an experimenter individually in the lab's testing room for a single session. Sessions with child participants took approximately 35-40 minutes and those with adult participants took approximately 25-30 minutes. Each session began with a general warm-up to make the participant comfortable, and a brief description of the upcoming tasks. Participants were told that they would watch a series of short puppet shows, and that the experimenter would have some brief questions for them following each show. The puppet shows were conducted first, starting with a training phase and then the experimental phase; the PPVT followed next for all participants, and the theory of mind measure was administered last (for child participants only).

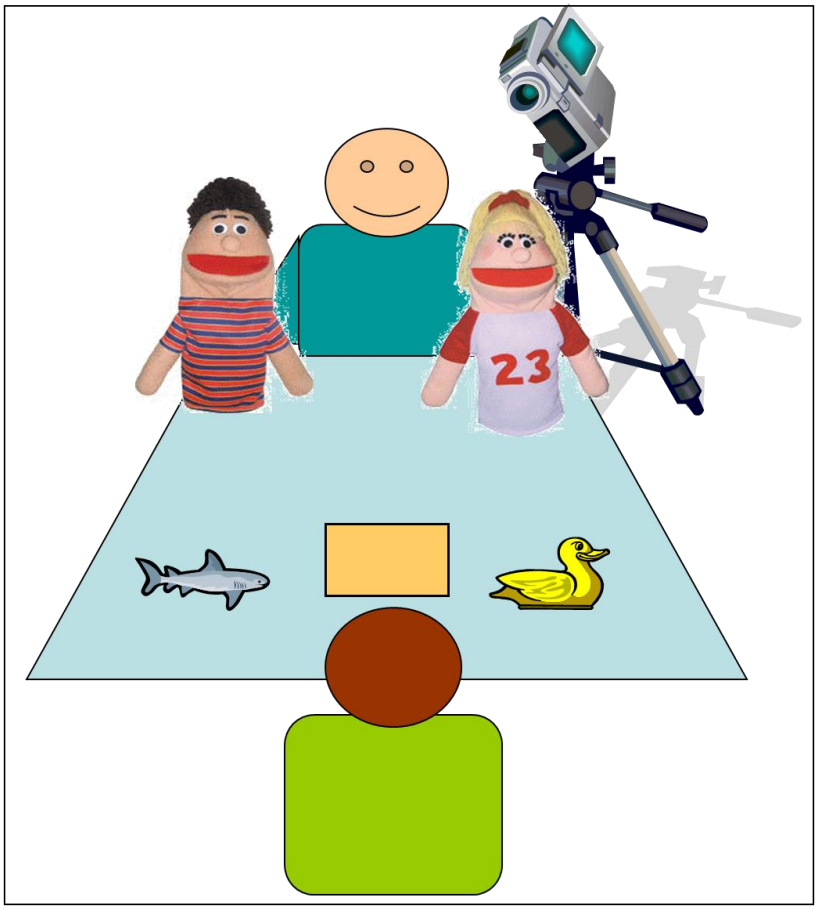
Training phase. First, participants were trained on the humour face scale to ensure they recognized the difference between scale points, and to ensure they had the opportunity to make use of all points on the scale. To train participants on the scale's use, the experimenter clearly labeled each face while pointing to it. Once each face was labeled the experimenter provided the participant with statements to give the participants an opportunity to practice using the scale. Each participant received a minimum of four practice trials, with corrective feedback.

Next, participants were trained on the use of the shark and duck props for ratings of speaker intent. The experimenter started by clearly labeling the duck, shark, and the answer box, and then placed each item in front of the participant in the position it would occupy during the testing phase. Participants were told that the duck is nice and friendly, and that the shark is mean and not friendly. They were told that after each story the experimenter would ask them a question about one of the story puppets and the

participant should show the experimenter their answer by placing the shark or the duck in the box. The participants were expected to place the duck in the answer box if they thought the speaker puppet was like the duck (i.e., was trying to be nice), and to place the shark in the answer box if they thought the speaker puppet was like the shark (i.e., was trying to be mean). Four practice trials were conducted in the training phase to ensure participants understood the appropriate use of these toys. Corrective feedback and praise were provided during the training phase.

Testing phase. For the experimental phase of the study participants were seated at a table, directly across from the experimenter. On each trial the answer box was positioned on the table, directly in front of the participant. On each side of the answer box was the shark or duck stuffed toy; one to the right and one to the left. The positions of these props were consistent within a participant, but were counterbalanced across participants. A schematic drawing of the physical set-up can be found in Figure 2. The narration of each puppet show was prerecorded and was played while the experimenter acted out the puppet show with the two story puppets and associated props. Following each puppet show, the experimenter posed four questions to the participant before proceeding to the next puppet show.

Figure 2. Testing Arrangement



The first question was used to tap into the participant's perception of speaker intent, using the shark/duck procedure. The experimenter asked "Was Shannon being like the shark or like the duck?" The order of the shark and duck labels was alternated across trials. After the participant made their selection and placed either the shark or the duck in the answer box, the experimenter moved on to the second question. The second question was used to determine if children recognize the speaker's true belief about the outcome of the situation: "When Shannon said *'that was a great play'* did she think the play was good or bad?" The presentation order of the good and bad options within the question was alternated across trials. The third question was used to investigate the participant's

perception of humorous intent via use of the face rating scale. The experimenter placed the scale in front of the participant and asked them to choose a face that reflected their assessment of the speaker's humorous intent: "Choose a face to show me how funny or serious Shannon was trying to be when she said *'that was a great play*."? The presentation order of the funny and serious labels within the question was alternated across trials. The final question was a manipulation check, used to assess whether the participant had paid attention to the relationship information presented at the start of the puppet show: "Were Shannon and Lisa sisters or had they just met?"

The entire session was videotaped by a camera, positioned over the experimenter's shoulder. The camera was angled to capture the participant's eye gaze during the shark/duck selection and to measure their response latency in choosing an object to place in the answer box. After the puppet shows all participants completed the PPVT-4. Finally, at the end of the session, child participants completed the theory of mind measure.

Coding. Responses to the speaker belief question were scored dichotomously, as either correct or incorrect. The correct assessment for literal criticisms and ironic criticisms was that the speaker believed the outcome was negative, and the correct assessment for the literal compliments was that the speaker believed the outcome was positive. Similarly, speaker intent judgments were coded dichotomously. The correct judgment for a literal or ironic criticism was that the speaker intended to be mean, like the shark, and for a literal compliment was that the speaker intended to be nice, like the duck. In addition to accuracy of intent judgments, video recording was used to assess children's eye gaze and response latencies for the decision making process. The digital

videos were examined on a frame-by-frame basis (1 frame = 33 ms) using FinalCut Pro 5.0.4, with audio and video signals fully synchronized. Response latencies were measured from the point when the experimenter began uttering the speaker intent question until the shark or the duck was placed in the answer box. As in Climie and Pexman (2008), the response latencies were further subdivided into three phases to capture early-, mid-, and late-stage processing. The early stage of processing spanned the time from the beginning of the experimenter's question until the participant initiated a reach towards an object; the middle stage of processing spanned the time from the initiation of movement until the participant made contact with either the shark or the duck; the later stage of processing spanned the time from when the participant touched the chosen response object until the object was placed in the answer box. Individual trials were excluded if the participant asked for clarification of the story details, if they asked the experimenter to repeat the prompt ("Was Shannon being like the shark or like the duck?"), if they provided a verbal response instead of a motor response, or if they directed their attention away from the testing set up (for instance, to tie their shoe, pick up a prop they dropped, etc.). Any of these interruptions would cause subsequent response latencies to be extraordinarily long and would cause the eye gaze to be uninformative so these trials were excluded before any analyses of response latency or eye gaze.

In keeping with the approach used in previous studies (Climie & Pexman, 2008; Dews et al., 1996; Pexman & Glenwright, 2007) humorous intent judgments were also scored dichotomously, with correct evaluations for the ironic criticisms being on the humorous side of the face scale (faces labeled very funny, funny or a little bit funny) and

correct evaluations for the literal criticisms and literal compliments being on the serious side of the scale (faces labeled very serious, serious or a little bit serious).

In order to be certain that participants were attending to the relationship information that was presented at the start of each trial, a manipulation check question was posed at the end of each trial. The experimenter asked the participant whether the characters in the preceding story had been siblings or strangers. If a participant responded incorrectly to the manipulation question on more than two trials their data was excluded from the analyses, but otherwise all trials were included. Adult participants and 7- to 8-year-old participants were very accurate in responding to the manipulation question and none of their data was excluded due to performance on the manipulation question. However, there were three 5- to 6-year-olds who responded incorrectly on the manipulation question on more than two trials; the data from these three participants were removed from all analyses.

A second coder examined 25% of the video data to ensure accuracy in measurement of the response latency and eye gaze data. The resulting Crohnbach's alphas demonstrated very good inter-rater reliability for the number of frames in each of the three response phases, with agreement from the start of the statement to the beginning of the reach being $\alpha = 1.00$, from the start of the reach until the participant touched the response object being $\alpha = 0.98$, and from touching the object to placing it in the answer box being $\alpha = 0.98$. Similarly, the alphas for eye gaze data showed high reliability between raters for the number of fixations to the shark ($\alpha = 0.97$) and to the duck ($\alpha = 0.94$), as well as for the duration of time spent looking at the shark ($\alpha = 0.94$) and the duck ($\alpha = 0.94$).

Design. Each participant viewed 10 puppet shows. Each show ended in a targeted remark made by the speaker puppet. On two trials the targeted remark was a literal compliment, on two it was a literal criticism, and on six it was an ironic criticism. The speaker puppet's remark was delivered using the intonation typically associated with that language type; literal remarks were delivered in a factual and sincere tone, and ironic remarks were delivered in a mocking and insincere tone. On half of the trials the speaker and target featured in the puppet show were siblings, and on the other half of trials the speaker and target were new acquaintances (i.e., they had just met during the activity that was being depicted in the puppet show). Across participants every puppet show was presented with every condition. As such, 10 versions of the materials were created, with counterbalanced order presentations within each version. All efforts were made to ensure an equal number of participants viewed each version.

Results

I will describe analyses of data from adult participants first, followed by analyses of children's data. There are two reasons for this delineation. First of all, one could argue that adults and children are expected to perform in qualitatively different ways, making a direct comparison between their data inappropriate. Second, although adult performance was measured to provide a baseline comparison for children's overall accuracy, response latency and eye gaze behavior, all hypotheses concerned children's performance, in particular, whether children used sibling relationship information as a cue to ironic intent, and how children's processing was impacted by that cue. Taken together, these points persuaded me that examination of adult and child data separately would be the most theoretically relevant and the most parsimonious way to address the questions of interest.

For both the adults and the children, results will be described first for the behavioral measures of irony appreciation, followed by explication of the response latency data, and finally an analysis of the eye gaze data. In addition, for the child data the correlations between the irony appreciation measures and the individual difference measures will be examined. Summary tables include descriptive data for all three language types, but for all analyses only the ironic criticism trials were included. This approach was taken for several reasons: 1) performance on literal trials lacked adequate variance for examination on several of the dependent measures, 2) because the design was a repeated-measures, any missing data on the literal trials caused the corresponding data on the ironic trials for those participants to be excluded from analyses, which increased the power issues related to having missing data, 3) understanding ironic language interpretation and processing was of greatest interest. This approach to the analyses means that for adults, the behavioral and processing data were analyzed using paired-samples *t*-tests to explore for differences between relationship types (sibling, stranger). For the child data, mixed-model ANOVA's were used to explore for differences between relationship type (a within-subjects factor), and age group (5- to 6-year-olds, 7- to 8-year-olds; a between-subjects factor).

Only subject-wise analyses were conducted, but Table 1 depicts the mean proportion correct for speaker intent judgments of ironic criticism trials for each item. This table illustrates that there was some variation in how successful participants were for each story, but with the exception of the "dessert" story, no single item was particularly problematic for adults or for children. The "dessert" story showed lower overall accuracy

for adults, but acceptable accuracy for children, so this story was kept in the analyses with all others.

Table 1

Participants' mean speaker intent accuracy on ironic criticism trials for each story (item)

Story	5-6 year olds	7-8 year olds	Adults
Soccer	.40	.67	.92
Nintendo	.50	.75	.85
Dessert	.63	.40	.42
Stampede	.65	.91	1.00
Shopping	1.00	.71	.91
Waterskiing	.63	.50	.90
Hide & Seek	.38	.33	.79
Bikes	.80	.88	1.00
Snowboarding	.83	.70	1.00
Painting	.33	.83	.92

Note. Includes correct speaker belief trials only.

Adult participants

Speaker belief. When the speaker made an ironic criticism the correct evaluation was that the speaker believed the outcome was bad. Adults' performance on the speaker belief question is summarized in Table 2.

Table 2

Adults' mean proportions correct on the speaker belief question

Ironic Criticism		Literal Criticism		Literal Compliment	
Sibling *n* = 20	Stranger *n* = 20	Sibling *n* = 20	Stranger *n* = 20	Sibling *n* = 20	Stranger *n* = 20
.93 (.18)	.98 (.11)	1.00 (.00)	.95 (.22)	1.00 (.00)	.95 (.23)

Note. Standard deviations in parentheses. For literal criticisms and ironic criticisms the correct evaluation was that the speaker believed the outcome was bad; for literal compliments the correct evaluation was that the speaker believed the outcome was good.

A paired-samples t-test for the ironic trials showed that there was no significant difference in speaker belief accuracy between the sibling condition (Mean proportion correct = .93, SD = .18) and the stranger condition (M = .98, SD = .11), $t(19) = 0.88$, p = .392. In addition, one-sample t-tests were conducted to test whether adults' speaker belief accuracy was significantly different from zero. Adults' speaker belief judgments differed from zero percent accuracy for both the sibling condition, $t(19) = 23.69$, $p < .001$, and the stranger condition, $t(19) = 39.00$, $p < .001$. Adults were highly accurate at determining that the ironic speaker held a belief that the outcome was a negative one, regardless of whether the speaker and target were siblings or strangers.

Speaker intent- aggression. A participant must understand that the ironic speaker holds a counterfactual belief about the event outcome in order to accurately determine the speaker's intent. Therefore, only trials for which there were correct evaluations on the speaker belief question were included in the speaker intent analyses. When the speaker made literal compliments the correct evaluation was that the speaker intended to be nice, whereas when the speaker made ironic criticisms and literal

criticisms the correct evaluation was that the speaker intended to be mean. The proportions of correct speaker intent judgments are listed in Table 3.

Table 3

Adults' mean proportions correct on the speaker intent question

Ironic Criticism		Literal Criticism		Literal Compliment	
Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
n = 20	*n* = 20	*n* = 20	*n* = 19	*n* = 20	*n* = 19
.87 (.20)	.88 (.21)	.85 (.37)	.79 (.42)	1.00 (.00)	1.00 (.00)

Note. Standard deviations in parentheses. For literal criticisms and ironic criticisms the correct evaluation was that the speaker intended to be mean; for literal compliments the correct evaluation was that the speaker intended to be nice.

A paired-samples *t*-test showed no significant difference between adults' accuracy at determining the speaker's ironic intent between the sibling condition (M = .87, SD = .20) and the stranger condition (M = .88, SD = .21), $t(19) = 0.13$, $p = .895$. In addition, one-sample *t*-tests conducted to test whether adults' speaker intent accuracy was significantly different from zero showed that their speaker intent judgments differed from zero percent accuracy for both the sibling condition, $t(19) = 19.51$, $p < .001$, and the stranger condition, $t(19) = 18.80$, $p < .001$.

Speaker intent- humour. Humorous intent is conceptualized as being a higher-order pragmatic goal, and children often have an understanding of ironic speakers' belief and intent to be aggressive before they understand the intent to be humorous (e.g., Pexman, Glenwright, Krol, & James, 2005). That is, children can comprehend a speaker's

counterfactual belief, as assessed by the speaker belief question, and that the speaker meant to be critical with their remark, before they further understand the speaker also intended to be simultaneously humorous. Therefore, only trials on which there were correct evaluations on the speaker belief question and the aggressive intent question were included in the speaker intent to be humorous analyses, for both adults' and children's data. When the speaker made an ironic criticism the correct evaluation was that the speaker intended to be humorous. Adults' accuracy on these humour intent judgments is listed in Table 4.

Table 4

Adults' mean proportions correct on the humour intent question

Ironic Criticism		Literal Criticism		Literal Compliment	
Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
n = 20	*n* = 20	*n* = 17	*n* = 15	*n* = 20	*n* = 19
.52 (.47)	.53 (.47)	1.00 (.00)	1.00 (.00)	.70 (.47)	.74 (.45)

Note. Includes correct trials only. Standard deviations in parentheses. For literal criticisms and literal compliments the correct evaluation was that the speaker intended the remark to be serious; for ironic criticisms the correct evaluation was that the speaker intended the remark to be funny.

The paired-samples *t*-test showed no significant effect of relationship, $t(19) = 0.11$, $p = .915$. Overall accuracy at detecting humorous intent was quite low, even for adults; humorous intent accuracy was no different between the sibling condition ($M = .52$, $SD = .47$) and the stranger condition ($M = .53$, $SD = .47$).

Response latencies. The response latency for each trial was recorded beginning with the start of the experimenter's question (e.g., "Was Shannon being like the duck or like the shark?") and ending when either the shark or the duck was placed inside the answer box. Response latencies were further subdivided into phases that represent early, middle or late-stage processing. The early stage included responding from the start of the question until the beginning of the participant's reach for either the shark or the duck; the middle stage included responding from the beginning of the reach movement until the participant actually touched the selected object; the late stage included the time that elapsed from when the participant touched the object until the object was placed inside the answer box. The response latencies reported here were only analyzed for trials in which the participant responded correctly to both the speaker belief and speaker intent questions. Three separate *t*-tests were conducted, one for the ironic trials at each response latency phase. Table 5 shows adults' mean response latencies for each phase of responding.

In the earliest phase of responding there was an effect of relationship that approached significance, $t(19) = 2.01$, $p = .058$, such that participants were faster to initiate movement towards the correct response object in the sibling condition (M = 2259.96 ms, SD = 679.44) than in the stranger condition (M = 2549.24, SD = 608.74). In the middle and late phases of processing there were no significant effects of relationship type: $t(19) = 1.05$, $p = .309$, and $t(19) = 1.35$, $p = .193$, respectively. Thus, adult participants did exhibit some sensitivity to the speaker relationship information in their early processing, as measured with response latency.

Table 5

Adults' mean response latencies (in milliseconds) during the early, middle, and late phases of responding

Response phase	Ironic Criticism		Literal Criticism		Literal Compliment	
	Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
	n = 20	*n* = 20	*n* = 16	*n* = 15	*n* = 20	*n* = 19
Early	2260 (679)	2549 (609)	3028 (1926)	2528 (950)	2185 (1044)	1959 (406)
Middle	668 (175)	629 (156)	615 (183)	590 (116)	680 (181)	643 (202)
Late	650 (129)	740 (289)	703 (131)	662 (149)	785 (163)	731 (209)
Total	3578 (663)	3918 (691)	4346 (1774)	3780 (896)	3650 (1035)	3333 (434)

Note. Standard deviations in parentheses. Includes correct trials only. Early response phase extends from the beginning of the statement "Was Shannon…." until the beginning of the participant's reach; middle phase extends from the beginning of the reach until the participant touches the response object; late phase extends from participant's touch of the response object until the object is placed in the answer box to end of the trial.

Eye gaze. Several components of eye gaze were of interest in the present study: first look accuracy, number of looks to the correct object, number of looks to the incorrect object, and proportion of total fixation to the correct object. Adult performance on each of these measures is summarized in Table 6.

First looks accuracy can provide insight into whether early processing is happening in a modular or parallel fashion. If, for instance, participants directed their first looks towards the duck on ironic criticism trials, this could reflect a tendency to consider a literal interpretation first. As such, in the case of first looks accuracy I tested for language effects rather than relationship effects. The adult participants had perfect accuracy on the literal criticism trials, so a paired-samples *t*-test was used to test for differences between the first looks accuracy for ironic criticisms and literal compliments. The analysis of first looks showed no significant difference in accuracy due to language type; adults were equally accurate in their first look to the correct response object for ironic criticisms ($M = .97$, $SD = .08$) and for literal compliments ($M = .97$, $SD = .11$), $t(18) = 0.25$, $p = .803$.

Table 6

Adults' first looks accuracy, numbers of fixations, and proportions of fixation duration

Eye gaze measure	Ironic Criticism		Literal Criticism		Literal Compliment	
	Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
	n = 18	*n* = 19	*n* = 11	*n* = 14	*n* = 17	*n* = 17
First looks accuracy (mean proportion)	.97 (.12)	.96 (.13)	1.00 (.00)	1.00 (.00)	1.00 (.00)	.94 (.25)
Number of fixations on correct object	1.04 (0.11)	1.00 (0.00)	1.18 (0.40)	1.07 (0.27)	1.00 (0.00)	1.00 (0.00)
Number of fixations on incorrect object	0.08 (0.26)	0.04 (0.13)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.06 (0.25)

Note. Includes correct trials only. Standard deviations in parentheses.

The remaining eye gaze analyses were conducted to test for differences between the relationship types. There was no variance in the number of fixations on the correct object for the stranger condition, so this variable could not be tested for adult participants. For the fixations to the incorrect object there was no difference between the sibling condition (M = .08, SD = .26) and stranger condition (M = .04, SD = .13), $t(17) = 0.52$, p = .612. In addition, the proportion of fixation duration to the correct object was not significantly different between the sibling condition (M = .95, SD = .17) and stranger condition (M = .96, SD = .15), $t(16) = 0.19$, p = .850. Taken together, these eye gaze data indicate that adults tend to direct their gaze towards the object that is correctly chosen (the shark) in the case of ironic criticisms. The presence of sibling relationship information does not have a significant impact on adults' overall consideration of the two response objects.

The proportions of looking time to the two response objects, the shark and the duck, were considered for the three separate response phases. These proportions are listed in Table 7.

Table 7

Adults' proportions of looking time to the correct and incorrect response objects during the early, middle and late phases of responding

Object	Ironic Criticism		Literal Criticism		Literal Compliment	
	Sibling *n* = 18	Stranger *n* = 19	Sibling *n* = 11	Stranger *n* = 14	Sibling *n* = 17	Stranger *n* = 17
			Early phase			
Correct	.11 (.11)	.09 (.11)	.13 (.14)	.09 (.10)	.06 (.06)	.03 (.03)
Incorrect	.02 (.08)	.01 (.03)	.03 (.08)	.02 (.06)	.00 (.00)	.00 (.00)
			Middle phase			
Correct	.67 (.32)	.65 (.29)	.62 (.41)	.76 (.35)	.87 (.24)	.79 (.32)
Incorrect	.02 (.08)	.02 (.07)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
			Late phase			
Correct	.05 (.07)	.06 (.08)	.04 (.08)	.03 (.05)	.04 (.08)	.04 (.08)
Incorrect	.01 (.01)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)

Note. Includes correct trials only. Proportions of looking time to the correct and incorrect object within each response phase do not sum to 1.0 because there were other possible looking locations, such as the response box and the experimenter. Those alternative looking locations were not analyzed here because looks to those locations were not relevant to the present study.

In the early phase of processing there was no difference between the proportion of fixation duration on the shark for the sibling condition (M = .11, SD = .11) and the stranger condition (M = .09, SD = .11), $t(17) = 0.70$, $p = .496$, and no difference in proportion of fixation duration to the duck between the sibling condition (M = .02, SD =

.08) and stranger condition (*M* = .01, *SD* = .03), *t*(17) = 0.66, *p* = .519. Similarly, for the middle phase of processing there were no differences between the sibling condition (*M* = .67, *SD* = .32) and the stranger condition (*M* = .65, *SD* = .29) for the fixation duration to the shark, *t*(17) = 0.36, *p* = .726, and no difference in proportion of fixation duration to the duck between the sibling condition (*M* = .02, *SD* = .08) and stranger condition (*M* =.02 , *SD* = .07), *t*(17) = 0.04, *p* = .966. Finally, in the late stage of processing analyses of looks to the duck were not conducted due to inadequate variance. In the late stage of processing there was no difference between the proportion of fixation duration on the shark for the sibling condition (*M* = .05, *SD* = .07) and the stranger condition (*M* = .06, *SD* = .08), *t*(17) = 0.16 , *p* = .872.

Child participants

Children's behavioral and processing data were first subjected to analyses to check for gender effects. One near-significant gender effect was found. Female participants were more accurate in their speaker belief assessments (Mean proportion correct = .56, *SD* = .37) than were male participants (*M* = .36, *SD* = .33), *t*(45) = 1.95, *p* = .057. No other gender effects were observed on any of the dependent measures. As such all subsequent analyses will be collapsed across gender for two reasons: 1) assessments of speaker intent are more important as a marker for irony understanding and there were no gender differences on that measure; 2) given the low statistical power afforded to some of the following analyses, it was not wise to further subdivide the data based on gender given that the gender effect was only observed for the measure of counterfactual belief understanding and not the measures of intent or processing.

Speaker belief. The proportions of correct speaker belief judgments made by children are listed in Table 8.

Table 8

Children's mean proportions correct on the speaker belief question

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
5-6 year olds	.39 (.38)	.37 (.40)	.91 (.28)	.93 (.27)	.85 (.36)	.85 (.36)
	n = 27	*n* = 27	*n* = 27	*n* = 27	*n* = 27	*n* = 27
7-8 year olds	.58 (.36)	.52 (.40)	.95 (.22)	1.00 (.00)	1.00 (.00)	.95 (.22)
	n = 20	*n* = 20	*n* = 20	*n* = 20	*n* = 20	*n* = 20

Note. Standard deviations in parentheses. For literal criticisms and ironic criticisms the correct evaluation was that the speaker believed the outcome was bad; for literal compliments the correct evaluation was that the speaker believed the outcome was good.

The speaker belief ANOVA showed that there was no interaction between relationship and age, $F(1, 45) = 0.25$, $p = .619$, and no main effect of relationship, $F(1, 45) = 1.22$, $p = .275$, or of age group, $F(1, 45) = 2.48$, $p = .122$. Children's overall accuracy on the speaker belief question was quite low (although in the range of that reported in previous studies, e.g., Climie & Pexman, 2008; Hancock et al., 2000); knowledge of a sibling relationship did not assist participants in making more accurate speaker belief judgments. However, a series of one-sample *t*-tests were conducted to test whether children's speaker belief accuracy was significantly different from zero. The *t*-tests for the younger group of children showed that speaker belief judgments differed

from zero percent accuracy for both the sibling condition, $t(26) = 5.41$, $p < .001$, and the stranger condition, $t(26) = 4.85$, $p < .001$. Likewise, the *t*-tests for the older group of children showed that speaker belief judgments also differed from zero percent accuracy for both the sibling condition, $t(19) = 7.28$, $p < .001$, and the stranger condition, $t(19) = 5.81$, $p < .001$. Thus, despite children's relatively low speaker belief accuracy their assessments differed from zero percent accuracy for sibling and stranger conditions.

Speaker intent- aggression. The proportions of correct speaker intent judgments made by children are listed in Table 9.

Table 9

Children's mean proportions correct on the speaker intent question

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
5-6 year olds	.68 (.39)	.73 (.34)	.96 (.20)	1.00 (.00)	.91 (.29)	1.00 (.00)
	n = 16	*n* = 16	*n* = 25	*n* = 25	*n* = 23	*n* = 23
7-8 year olds	.80 (.31)	.63 (.44)	.95 (.23)	.90 (.31)	.95 (.22)	.95 (.23)
	n = 17	*n* = 14	*n* = 19	*n* = 20	*n* = 20	*n* = 19

Note. Standard deviations in parentheses. For literal criticisms and ironic criticisms the correct evaluation was that the speaker intended to be mean; for literal compliments the correct evaluation was that the speaker intended to be nice.

These accuracy results only included trials for which the participant understood the speaker's belief. Results showed that the Relationship x Age Group interaction was not significant, $F(1, 26) = 1.04$, $p = .318$; likewise, there were no main effects of

relationship, $F(1, 26) = 0.49$, $p = .491$, or of age group, $F(1, 26) = 0.02$, $p = .889$. A series of one-sample *t*-tests were conducted to test whether children's speaker intent accuracy was significantly different from zero. The *t*-tests for the younger group of children showed that speaker intent judgments differed from zero percent accuracy for both the sibling condition, $t(15) = 6.92$, $p < .001$, and the stranger condition, $t(15) = 8.50$, $p < .001$. Likewise, the *t*-tests for the older group of children showed that speaker intent judgments differed from zero percent accuracy for both the sibling condition, $t(16) = 10.78$, $p < .001$, and the stranger condition, $t(13) = 5.32$, $p < .001$.

Speaker intent- humour. The proportions of children's correct judgments of speaker intent to be humorous are listed in Table 10, and include only cases where the participant correctly identified the speaker's belief and intent.

Table 10

Children's mean proportions correct on the humorous intent question

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
5-6 year olds	.28 (.45)	.33 (.43)	.79 (.41)	.84 (.37)	.76 (.44)	.61 (.50)
	n = 13	*n* = 14	*n* = 24	*n* = 25	*n* = 21	*n* = 23
7-8 year olds	.63 (.50)	.50 (.53)	.78 (.43)	.83 (.38)	.84 (.37)	.79 (.42)
	n = 16	*n* = 10	*n* = 18	*n* = 18	*n* = 19	*n* = 19

Note. Includes correct trials only. Standard deviations in parentheses. For literal criticisms and literal compliments the correct evaluation was that the speaker intended the remark to be serious; for ironic criticisms the correct evaluation was that the speaker intended the remark to be funny.

Results showed that the Relationship x Age Group interaction was not significant, $F(1, 19) = 1.79$, $p = .197$; likewise, there was no main effect of relationship, $F(1, 19) = 1.79$, $p = .197$, or of age, $F(1, 19) = 0.80$, $p = .384$.

Response latencies. Table 11 depicts the mean response latencies for each response phase (early, middle and late) for the younger and the older child groups. Analysis of the latencies for ironic criticisms showed that in the early stage of processing, there were no differences in response latency due to interactions between relationship and age group, $F(1, 19) = 0.61$, $p = .444$, or due to a main effect of age group $F(1, 19) = 0.90$, $p = .354$.

Table 11

Children's mean response latencies (in milliseconds) during the early, middle and late phases of responding

Age Group	Ironic Criticism		Literal Criticism		Literal Compliment	
	Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
			Early phase			
5-6 year olds	2536 (1895)	3093 (2532)	1754 (1347)	2366 (1843)	2041 (1688)	1581 (1772)
	n = 13	*n* = 14	*n* = 20	*n* = 24	*n* = 19	*n* = 22
7-8 year olds	1928 (1317)	2572 (2788)	1641 (1729)	1593 (969)	2090 (1369)	3098 (2590)
	n = 16	*n* = 10	*n* = 15	*n* = 14	*n* = 18	*n* = 17
			Middle phase			
5-6 year olds	1139 (1366)	876 (386)	543 (182)	617 (261)	597 (312)	696 (629)
	n = 13	*n* = 14	*n* = 20	*n* = 24	*n* = 19	*n* = 22
7-8 year olds	529 (115)	675 (196)	664 (229)	566 (107)	598 (143)	557 (129)
	n = 16	*n* = 10	*n* = 15	*n* = 14	*n* = 18	*n* = 17
			Late phase			
5-6 year olds	1109 (282)	1639 (1315)	998 (596)	1401 (1772)	787 (336)	906 (451)
	n = 13	*n* = 14	*n* = 20	*n* = 24	*n* = 19	*n* = 22
7-8 year olds	1079 (641)	1145 (465)	948 (398)	854 (684)	997 (623)	893 (289)
	n = 16	*n* = 10	*n* = 15	*n* = 14	*n* = 18	*n* = 17
			Total response time			
5-6 year-olds	4784 (2248)	5609 (3689)	3203 (1443)	4441 (2885)	3425 (1736)	3183 (1789)
	n = 13	*n* = 14	*n* = 20	*n* = 24	*n* = 19	*n* = 22
7-8 year-olds	3537 (1612)	4392 (2319)	3254 (1936)	3036 (739)	3685 (1492)	4548 (2583)
	n = 16	*n* = 10	*n* = 15	*n* = 14	*n* = 18	*n* = 17

Note. Standard deviations in parentheses. Includes correct trials only.

However, the effect of relationship approached significance[1], $F(1, 19) = 3.89$, $p = .063$, where response latencies were faster for the sibling condition ($M = 2041.54$ ms, $SD = 1758.36$) than for the stranger condition ($M = 2858.95$ ms, $SD = 2761.52$). During the middle phase of processing there were no differences in response latency due to interactions between relationship and age group, $F(1, 19) = 0.92$, $p = .349$, or due to main effects of relationship $F(1, 19) = 0.21$, $p = .650$, or age group $F(1, 19) = 2.64$, $p = .121$. Similarly, at the final stage of processing, there were no differences in response latency due to interactions between relationship and age group, $F(1, 19) = 0.61$, $p = .446$, or due to main effects of relationship $F(1, 19) = 1.72$, $p = .205$, or age group $F(1, 19) = 1.10$, $p = .308$. As with the behavioral measures, children in both the younger age group and the older age group demonstrated little sensitivity to the speaker cue of sibling relationship. The near-significant effect that was observed occurred in the early phase of processing and indicates that children may have some sensitivity to the sibling relationship cue early in irony interpretation.

Eye gaze. Eye gaze measures included the proportion of first looks directed towards the correct response object, the number of fixations on the correct object, the number of fixations on the incorrect object, and the proportion of the total fixation duration to the correct object. Only trials for which there was at least one look to the shark or the duck were included in the analyses of eye gaze measures. That is, if

[1] This effect mirrors the one seen amongst adult participants for the early stage of processing. I conducted an additional analysis, a 2 (relationship: sibling, stranger) x 3 (age group: 5-6-year-olds, 7-8-year-olds, adults) mixed model ANOVA, to determine if the main effect of relationship was significant with all three age groups included in the analysis. Results of this analysis showed there was no interaction of age group and relationship, and no main effect of age group. The main effect of relationship did, however, reach significance, such that responses were faster in the sibling condition ($M = 2148.08$ ms, $SD = 1333.19$) than in the stranger condition ($M = 2707.87$, $SD = 2003.39$), $F(1, 38) = 7.66$, $p = .009$, $\eta_p^2 = .168$, when all 3 age groups were included. This analysis indicates that increased statistical power could reveal reliable use of this cue in both children and adults.

participants kept their gaze fixated on the experimenter and did not look to either response object, then they did not contribute any eye gaze data on that trial.

First looks. The proportions of first looks that were directed towards the correct response object (i.e., the shark) are listed in Table 12.

Table 12

Children's mean first looks accuracy

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
5-6 year olds	.86 (.32)	.89 (.20)	1.00 (.00)	.91 (.29)	.88 (.33)	1.00 (.00)
	n = 11	*n* = 12	*n* = 18	*n* = 22	*n* = 17	*n* = 20
7-8 year olds	.94 (.16)	.83 (.35)	.77 (.44)	.93 (.27)	.94 (.25)	.81 (.40)
	n = 13	*n* = 9	*n* = 13	*n* = 14	*n* = 17	*n* = 16

Note. Includes correct trials only.

These looks were analyzed to determine whether eye gaze reflected consideration of the correct response object from the earliest measured time point in processing, the first fixation. There were no significant differences in the proportions of first looks directed at the correct object for the language and age interaction, $F(2, 42) = 0.13$, $p = .191$, or for the main effects of language, $F(2, 42) = 1.72$, $p = .191$, or age group $F(1, 21) = 0.15$, $p = .700$. Older and younger children were equally accurate at directing their first look towards the correct response object, regardless of language type. This indicates that in the ironic criticism trials, children were as likely to consider an ironic interpretation

first for the ironic criticism trials as they were to consider a literal interpretation first on the literal trials.

Number of looks. The mean numbers of fixations to the correct and incorrect response objects for correct trials are listed in Table 13. There were no significant differences in the numbers of fixations to the correct object (the shark) for the Relationship x Age interaction, $F(1, 14) = 1.51$, $p = .239$, or for the main effects of relationship, $F(1, 14) = 3.41$, $p = .086$, or of age, $F(1, 14) = 2.15$, $p = .164$. Similarly, the analysis of the number of fixations to the incorrect object (the duck) demonstrated no Relationship x Age interaction, $F(1, 14) = 0.16$, $p = .697$, and no main effects of relationship, $F(1, 14) = 1.44$, $p = .250$, or of age, $F(1, 14) = 0.33$, $p = .576$.

Table 13

Children's mean numbers of fixations on the correct response object and the incorrect response object

Age Group	Ironic Criticism		Literal Criticism		Literal Compliment	
	Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
			Fixations on Correct Object			
5-6 year olds	1.05 (0.47)	1.46 (0.81)	1.11 (0.32)	1.09 (0.43)	1.06 (0.24)	1.15 (0.37)
	n = 11	*n* = 12	*n* = 18	*n* = 22	*n* = 17	*n* = 20
7-8 year olds	0.96 (0.14)	1.28 (0.67)	1.23 (0.60)	1.14 (0.36)	1.00 (0.00)	1.07 (0.26)
	n = 13	*n* = 9	*n* = 13	*n* = 14	*n* = 17	*n* = 16
			Fixations on Incorrect Object			
5-6 year olds	0.26 (0.40)	0.31 (0.41)	0.06 (0.25)	0.27 (0.46)	0.18 (0.39)	.00 (.00)
	n = 11	*n* = 12	*n* = 18	*n* = 22	*n* = 17	*n* = 20
7-8 year olds	0.06 (0.16)	0.28 (0.44)	0.23 (0.44)	0.07 (0.27)	0.06 (0.25)	0.27 (0.59)
	n = 13	*n* = 9	*n* = 13	*n* = 14	*n* = 17	*n* = 16

Note. Standard deviations in parentheses. Includes correct trials only.

Proportion of looking time to target objects during each response phase. The proportions of time that participants spent looking at the shark and the duck during the early, middle and late phases of responding are listed in Table 14.

Analyses were conducted on the proportions of looking time to the two response objects (the shark and the duck) for each phase of responding. In the prior analyses reported, the first looks accuracy, the number of looks to the shark and duck, and the proportion of looking time to the correct object were computed *only* including trials for which there were at least one look to either the shark or the duck. This was done to examine the relative consideration given to each of the two response objects. In the results reported in this section, however, the analyses were conducted on proportions of looks to the shark and duck out of all potential looking locations (i.e., the shark, duck, experimenter, answer box, participant's hands, and story puppets and props).

Table 14

Children's proportions of looking time to the correct and incorrect response objects during the early, middle and late phases of responding

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Sibling	Stranger	Sibling	Stranger	Sibling	Stranger
			Early phase			
5-6 year olds						
Correct	.09 (.11)	.10 (.12)	.08 (.11)	.15 (.15)	.12 (.10)	.15 (.17)
Incorrect	.02 (.06)	.06 (.13)	.03 (.11)	.02 (.05)	.02 (.06)	.01 (.04)
	n = 11	*n* = 12	*n* = 18	*n* = 22	*n* = 17	*n* = 20
7-8 year olds						
Correct	.12 (.13)	.19 (.27)	.16 (.14)	.15 (.18)	.08 (.11)	.11 (.13)
Incorrect	.01 (.03)	.10 (.26)	.03 (.06)	.00 (.00)	.00 (.00)	.04 (.08)
	n = 13	*n* = 9	*n* = 13	*n* = 14	*n* = 17	*n* = 16
			Middle phase			
5-6 year olds						
Correct	.70 (.39)	.58 (.35)	.90 (.21)	.76 (.31)	.79 (.29)	.82 (.28)
Incorrect	.11 (.25)	.02 (.09)	.00 (.00)	.01 (.01)	.00 (.00)	.00 (.00)
	n = 11	*n* = 12	*n* = 18	*n* = 22	*n* = 17	*n* = 20
7-8 year olds						
Correct	.75 (.21)	.65 (.38)	.88 (.21)	.87 (.10)	.82 (.25)	.72 (.35)
Incorrect	.03 (.12)	.08 (.26)	.00 (.00)	.00 (.00)	.00 (.00)	.05 (.21)
	n = 13	*n* = 9	*n* = 13	*n* = 14	*n* = 17	*n* = 16
			Late phase			
5-6 year olds						
Correct	.24 (.29)	.19 (.29)	.07 (.09)	.14 (.21)	.18 (.28)	.18 (.26)
Incorrect	.00 (.00)	.01 (.02)	.00 (.00)	.04 (.12)	.00 (.00)	.00 (.00)
	n = 11	*n* = 12	*n* = 18	*n* = 22	*n* = 17	*n* = 20
7-8 year olds						
Correct	.14 (.24)	.21 (.28)	.08 (.14)	.15 (.32)	.13 (.25)	.14 (.28)
Incorrect	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
	n = 13	*n* = 9	*n* = 13	*n* = 14	*n* = 17	*n* = 16

Note. Standard deviations in parentheses. Includes correct trials only. Proportions of looking time to the correct and incorrect object within each response phase do not sum to 1.0 because there were other possible looking locations, such as the response box and the experimenter. Those alternative looking locations were not analyzed here because looks to those locations were not relevant to the present study.

In the early phase of processing there were no significant differences in the proportion of looking time to the shark based on an interaction between relationship and age, $F(1, 17) = 0.26$, $p = .620$, or based on main effects of relationship type, $F(1, 17) = 0.52$, $p = .481$, or of age group, $F(1, 17) = 0.71$, $p = .410$. Similarly, in the early phase of processing there were no significant differences in the proportion of looking time to the duck based on an interaction between relationship and age, $F(1, 17) = 0.62$, $p = .442$, or based on main effects of relationship type, $F(1, 17) = 1.20$, $p = .288$, or of age group, $F(1, 17) = 0.42$, $p = .527$.

In the middle phase of processing there were no significant differences in the proportion of looking time to the shark based on an interaction between relationship and age, $F(1, 17) = 0.01$, $p = .912$, or based on main effects of relationship type, $F(1, 17) = 1.86$, $p = .191$, or of age group, $F(1, 17) = 0.90$, $p = .356$. Similarly, in the middle phase of processing there were no significant differences in the proportion of looking time to the duck based on an interaction between relationship and age, $F(1, 17) = 0.97$, $p = .340$, or based on main effects of relationship type, $F(1, 17) = 0.16$, $p = .695$, or of age group, $F(1, 17) = 0.02$, $p = .881$.

In the final phase of processing there was insufficient variance in several cells to permit analysis of looks to the duck. For looks to the shark there were no significant differences in the proportion of looking time based on an interaction between relationship and age, $F(1, 17) = 0.18$, $p = .674$, or based on main effects of relationship type, $F(1, 17) = 0.02$, $p = .896$, or of age group, $F(1, 17) = 0.10$, $p = .758$.

Individual difference measures. Partial correlations were conducted for the child participants as a group (controlling for age), for ironic trials only (Table 15). Correlations were conducted to examine relationships between each of the above-reported dependent measures and each of the individual difference measures: theory of mind, PPVT raw score, CCC pragmatic score, SRQ conflict score, SRQ status score, SRQ warmth score, and number of siblings in family (from the demographic questionnaire).

Three significant correlations emerged, and one additional correlation was near-significant. First, response latency during the middle response phase (from the initiation of movement until the participant touched the response object) was positively correlated with performance on theory of mind, $r(29) = .46$, $p = .009$. Children with a better developed theory of mind actually took more time to reach out and touch the shark. Next, the relationship between theory of mind and the proportion of looking time to the shark during the middle phase of processing approached significance, $r(26) = -.37$, $p = .054$. Participants with better developed theory of mind spent less time directing their gaze to the shark during the second phase of processing, despite the fact that these same participants spent more relative time in that response phase according to their response latencies. Finally, two aspects of the sibling relationship questionnaire were related to the proportion of looks directed at the duck (i.e., the incorrect object), during processing of the ironic utterance. The first of these was a relationship between the conflict subscale of the SRQ and the proportion of looks to the duck during the early phase of processing, $r(26) = .41$, $p = .031$, indicating the children who have a higher degree of conflict in their own sibling relationship spend a greater proportion of time considering the duck as a response alternative in the early phase of processing. The second of the correlations with

the sibling relationship questionnaire was a correlation between the status subscale of the SRQ and the proportion of looks to the duck during the last phase of processing, $r(26) = .42$, $p = .027$, indicating that children who have more status in their own sibling relationship spend a greater proportion of time considering the duck as a response alternative in the late phase of processing. These final two correlations involving the proportion of looks to the duck ought to be interpreted with caution, however, since the overall proportion of looks to the duck at either of these stages was quite low and created a restriction of range that may be problematic for the interpretation of correlations. Furthermore, given the large number of correlations that were conducted, it is possible that the effects observed could be spurious, so interpretation of the reported correlations should be made with caution.

Table 15

Partial correlations between children's performance on irony task measures and individual difference measures

Dependent measure		ToM	PPVT	CCC Pragmatic	SRQ Conflict	SRQ Warmth	SRQ Status	# of siblings
Speaker Belief	*n* = 44	-.01	.07	.06	-.03	.05	-.10	-.01
Speaker Intent	*n* = 32	-.07	-.24	.12	-.19	.06	-.10	.13
Humour Intent	*n* = 26	.13	.26	.30	.13	-.07	.21	.15
Latency-early	*n* = 26	-.01	.05	.20	.02	.12	.04	.03
Latency- middle	*n* = 26	.46*	.01	.02	-.23	.11	-.26	.15
Latency- late	*n* = 26	-.02	-.07	.22	-.25	.24	-.17	-.03
First look correct	*n* = 25	.17	-.18	-.02	-.02	-.22	.07	.28
Looks to correct	*n* = 25	.20	-.02	.22	.04	.28	-.07	.03
Looks to incorrect	*n* = 25	.23	-.07	.26	.16	.16	-.32	-.26
Proportion time to correct	*n* = 25	-.09	.09	-.18	-.33	-.18	.33	.32
Shark proportion-early	*n* = 25	-.02	.03	.05	.11	.25	.11	-.13
Duck proportion- early	*n* = 25	.07	.03	.20	.41*	<.01	-.05	-.28
Shark proportion- middle	*n* = 25	-.37*	.07	-.15	-.14	-.22	.32	-.08
Duck proportion- middle	*n* = 25	.04	-.05	.05	.26	.15	-.20	-.19
Shark proportion- late	*n* = 25	.14	.15	-.10	-.31	.12	-.08	-.03
Duck proportion- late	*n* = 25	-.10	.04	.02	-.07	.11	.42*	.10

Note. Correlations reported here only include data from child participants (younger and older child groups combined) and ironic trials. Age was partialled-out of these correlations. * $p < .05$

Discussion

The purpose of Experiment 1 was to determine a) whether knowledge of a speaker-target sibling relationship would facilitate irony appreciation for 5- to 8-year-old children; b) how processing was impacted by the presence of the sibling relationship information; and c) whether individual difference measures like theory of mind, language proficiency, and sibling relationship quality were related to overall irony appreciation.

Adult participants were included to establish a baseline for the behavioral and processing measures. Not surprisingly, adult performance on the speaker belief and speaker intent questions was quite high; adults were adept at understanding that a speaker meant to communicate ironic intent. Knowledge that the story characters were siblings or strangers did not have any bearing on adults' accuracy with these measures. Only one of the processing measures showed any sensitivity to the speaker relationship cue: adults tended to be faster to initiate movement to select the shark in the ironic criticism trials when the speaker and target were siblings than when they were strangers. For adults it appears that sibling relationships may be more congruent with ironic language use than are newly-formed relationships. Although no predictions were made regarding adult performance, this finding does demonstrate that the sibling relationship information can be integrated and used in the processing of ironic utterances. Of greater interest, of course, was whether children were sensitive to sibling relationship information, and whether there were developmental improvements in comprehension and processing.

First, with respect to the behavioral measures, it was predicted that the older group of children would outperform the younger group of children on comprehension of ironic criticisms for each of the three comprehension measures. For speaker belief,

speaker intent, and humorous intent there was no age-related improvement among children for accuracy in responding to ironic criticisms. It was also predicted that performance on these three comprehension measures would be greater for the sibling condition than for the stranger condition. This prediction was not upheld; sibling relationship information did not improve children's comprehension on the speaker belief, speaker intent, or humorous intent questions.

With respect to response latencies it was predicted that older children would make correct responses more quickly than would younger children. This prediction was not supported. At all three response phases, the response latencies were equivalent for both age groups. It was also predicted that children's response latencies would be faster for the sibling condition than for the stranger condition. There was limited evidence to support this prediction. However, children did display the same tendency as adults in the earliest stage of processing, whereby they were faster to initiate movement to the shark (indicating an ironic interpretation of the utterance) when the characters were siblings than when they were strangers. This finding warrants cautious interpretation since it was not actually significant. However, the trend approached significance and points to a possible sensitivity to the sibling relationship cue. In addition, when the adult and child data were combined in a supplementary analysis (see footnote 1), the main effect of relationship on early response latency did reach significance.

Several predictions were made about eye gaze behavior. First, it was predicted that the older group of children would be more accurate in directing their first look towards the shark on the ironic trials than would the younger children. This prediction was not upheld. Children in both age groups were highly accurate in the first looks

measure and there was no developmental improvement observed. The prediction regarding first looks accuracy was that lower accuracy in the ironic trials would support literal-first interpretations of processing but equal looks across all language types would support interactive interpretations of processing. The first looks data showed equal looks across all language types, supporting the interactive accounts. Based on their first gaze fixation, children were equally as likely to consider an ironic interpretation on ironic trials as they were to consider a literal interpretation on literal trials. Next, it was predicted that the mean number of fixations to the correct object in ironic trials would be higher for the stranger condition than for the sibling condition. There was no support for this prediction. Finally, it was predicted that the proportion of fixation duration on the correct object would be higher in the sibling condition than the stranger condition; this prediction was also unsupported. Across all eye gaze measures, there was no evidence that sibling relationship information was being used in consideration of the response objects and in decision making.

Children did not show evidence of using sibling relationship information in their assessments of speaker belief and intent, or in their eye gaze, and showed a modest sensitivity to the cue in their response latencies. There are several reasons why children in the present experiment may not have demonstrated sensitivity to the sibling relationship cue. It is possible that relationship information was not additionally helpful over and above the other cues present: context incongruity and intonation. Second, it is possible that sibling information was not manipulated in a meaningful or realistic manner. The scenarios depicted in the study may not be representative of the sorts of conversational contexts the participants face with their siblings. Third, the participant

sample had widely variable family structures; some children had siblings who were much older or much younger and some children had one sibling while others had many. Despite the fact that all participants ought to have had a sense of how sibling relationships operate, it might be the case that some participants had inadequate experience with their own siblings, if there was a large age difference, to understand the nuances of typical ironic communication amongst siblings. Finally, it is very likely that issues of statistical power limited the detection of true effects. A power analysis demonstrated that the present study included sufficient participants to detect effects that were moderate in size. However, if the effect size of a sibling relationship effect is actually small, the present study would not have adequate power to detect those smaller effects. Moreover, the number of participants was sufficient at the outset, but after each successive behavioral analysis there were filters applied to ensure that only cases for which participants correctly addressed the prior questions were included. With only modest accuracy on those measures I used as filters, many fewer cases remained after incorrect responses to the speaker belief question and speaker intent questions were discarded. Statistical sensitivity was further attenuated by missing data for the processing measures on trials where participants provided verbal responses, asked for clarification, played with the props, or took an extraordinarily long time to provide a response. The remaining sample size on the processing measures was thus quite small. In the early phase of processing both adults and children tended to respond more quickly in the sibling condition than in the stranger condition, but these effects were non-significant. When their data was combined, the improvement in statistical power allowed the effect to reach significance,

further supporting the notion that statistical power may be a limitation to the analyses in the present study.

Several individual differences measures were examined for correlations with the dependent measures from the irony appreciation task. Theory of mind, receptive vocabulary (PPVT), pragmatic language understanding (CCC), sibling relationship quality (SRQ), and number of siblings were all considered. Very few correlations emerged. Relationships were observed in the middle phase of processing between theory of mind and response latency, and theory of mind and looks to the shark. These correlations indicated that children with higher theory of mind scores took longer to respond and also spent less time looking at the shark during this phase. This finding is counterintuitive to expectations since a more well-formed theory of mind ought to facilitate, rather than inhibit, responding. It may be the case that children with better developed theory of mind were taking more time in that processing phase to consider alternatives carefully from different perspectives.

This experiment can be viewed as a first step towards an examination of the types of sibling relationships that facilitate understanding of ironic language. It is apparent from this study that not all sibling contexts might be sufficient as a cue to ironic intent. The results point to several recommendations for future research on this topic. First, better control of participant characteristics might allow small but important effects to emerge. For instance, family size, birth order, age spacing, and gender composition could all be controlled for more specific examination. Second, the characters used in the puppet shows did not always have features that supported the notion that they were siblings. Some puppet pairs had different hair and skin colors, and vastly different hair type and

texture, which diminished their similarity. This may have led to some participants not viewing the characters as being siblings, despite the participants' ability to remember the 'sibling' label.

Although Experiment 1 provided only limited evidence that a sibling relationship can cue ironic intent, there are several other speaker traits that may be relevant cues, and I examined one of those in Experiment 2: the speaker's similarity to the listener.

Experiment 2

One reason that children find irony to be a very critical (and not humorous) way of speaking may be because they readily identify with the target of a remark and empathize with the potential scorn (Pexman et al., 2005). This tendency to identify with the target may prevent children from appreciating the opportunity for humour afforded by ironic remarks. In Experiment 2 I will test the possibility that if children identify strongly with the speaker of ironic remarks they may have greater accuracy of speaker intent assessments, including perception of the speaker's intent to be humorous.

The perception of ironic intent can depend on the perspective taken (Pexman & Olineck, 2002a; Toplak & Katz, 2000) and children are likely to identify with the target of a remark (Harris & Pexman, 2003; Pexman et al., 2005). Children's view that irony is aggressive can be linked to the fact that the use of ironic language can violate politeness norms. Irony can be face-threatening for both the speaker and the target (Brown & Levinson, 1987). Children's sensitivity to this notion may prevent them from seeing the opposite viewpoint: that ironic language can be used to *save* face by veiling a criticism in positively-worded language, such as when ironic criticisms are employed (Brown & Levinson, 1987). Even though children likely identify with the target of a critical remark,

we know that children do use irony voluntarily, and do so with the apparent intent of joking and teasing, or being something other than strictly aggressive (Pexman, Zdrazilova, McConnachie, Deater-Deckard, & Petrill, 2009; Whalen & Pexman, 2010). When children themselves use ironic language their own intent may be clear to them. That is, children likely view themselves as nice, and recognize their own jovial or light-hearted intent. However, children in early and middle childhood may still struggle with the challenges of considering the intent from the speaker's perspective.

Children use speaker trait information when evaluating ironic remarks made by others. That is, children judge critical remarks made by ‘nice’ speakers as less aggressive than the same remarks made by ‘mean’ speakers (Pexman et al., 2006). If the speaker puppet is one that the child has established is 'like them' then the child may be able to more readily gain insight into the speaker’s attitude and may recognize that a speaker's intent with an ironic remark can be positive (or, at least, not strictly negative).

Past studies involving puppet similarity manipulations have indicated that preschoolers are sensitive to perceived similarity (see Reyes-Jacques & Echols, 2013, for a review). Reyes-Jacques and Echols (2013) introduced preschool children to puppets by providing descriptions that would make the puppets either appear to be similar to the participant or dissimilar from the participant. The ‘similar’ puppet shared the same hair color and favorite food as the participant and the ‘dissimilar’ puppet had a different hair color and preferred the participant’s least-favorite food. Children were then expected to choose one puppet as an informant in helping them identify the name of a novel object. Reyes-Jacques and Echols found that 5-year-olds chose the similar puppet to provide a label for a novel object more than they chose the dissimilar puppet to provide a label.

Further to this, when familiar and unfamiliar puppets were used in a labeling task, 3-year-olds did not choose the similar puppet any more often than the dissimilar puppet when they were both familiar. However, when two familiar puppets were used, the 5-year-olds prioritized requests of a similar puppet over those of a dissimilar puppet. This evidence indicates that perceived similarity to a puppet can be induced amongst young children in a laboratory setting, and children use this similarity information when making judgments about the usefulness of a knowledge source. If young children can use similarity information to judge the suitability of an epistemic source, then they may also use similarity information to adopt the other's perspective more readily.

Perspective taking is a critical social and cognitive skill. Perspective taking is closely related to theory of mind, and refers generally to the child's ability to recognize that other agents have thoughts, feelings, and access to information sources that differ from the child's own (e.g., Astington, 1998). The development of theory of mind extends into middle childhood (and possibly beyond), but by age 5 typically developing children generally have the ability to view situations from the perspective of another person. In fact, some sources have demonstrated that children can judge others' knowledge states as early as age 2-3 (O'Neill, 1996; Perner & Leekham, 1986). Theory of mind allows children to understand that speakers may intend something very different from the spoken message. The purpose of the present study was not to address whether children have the ability to take another's perspective, but instead to determine if children would appreciate ambiguous language (in the form of irony) differently if they were encouraged to take the perspective of the speaker.

Neuroimaging evidence shows that the ventromedial prefrontal cortex (vMPFC) is activated when children take the perspective of another person, just as it is when they think about themselves being involved in a scenario (Ames, Jenkins, Banaji, & Mitchell, 2008). However, the vMPFC does not engage in the same way when children consider a scenario from a third-person perspective. If children are able to adopt another's perspective in situations where ironic language is used, they may be able to see beyond the tendency to identify with the target of a critical ironic remark. Identifying with another person may encourage the child to take their perspective; it may facilitate the adoption of common ground, which is considered to be a necessary condition for irony interpretation (Clark, 1992). The aim in the proposed Experiment 2 was to determine if children can more readily detect and process ironic remarks when they are encouraged to take another's perspective.

To accomplish this, all testing sessions in Experiment 2 began with the participant choosing one puppet that was 'like them' (the 'similar' condition) and one puppet that was 'not like them' (the 'dissimilar' condition) from two options presented by the experimenter. The chosen puppets were then each used as speakers in half of the puppet shows the participant watched during the testing procedure. By this method, I hoped to determine whether children's impressions of ironic remarks would vary as a function of their identification with a speaker. In other respects, the stories used in Experiment 2 were very similar to those described for Experiment 1. Sample puppet show stories can be found in Appendix A.

Predictions

1. Children in the older age group (7- to 8-year-olds) will demonstrate greater accuracy for ironic criticism understanding than will children in the younger age group (5- to 6-year-olds). This effect was anticipated for assessments of speaker belief, speaker intent, and humorous intent.
2. If similarity is a cue to speaker intent for irony, then overall accuracy will be higher on each of the behavioral measures of irony appreciation for the similar conditions than for the dissimilar conditions.
3. Children in the older group will have faster response latencies to ironic criticisms than will children in the younger group.
4. If similarity is a cue to speaker intent for irony, then response latencies to ironic criticisms in the similar conditions will be faster than those in the dissimilar conditions.
5. The accuracy of the first fixation during ironic trials will be higher for children in the older group than for children in the younger group.
6. If the literal interpretation is considered more often early in processing, as modular processing accounts would predict, then the accuracy of the first fixation to the correct object during ironic trials will be lower than that during the literal trials. However, if the ironic interpretation is considered as readily as the literal interpretation early in processing, as interactive processing accounts would predict, then the accuracy of the first fixation during ironic trials and literal trials will be equal.

7. If similarity is a cue to speaker intent for irony, then the mean number of fixations to the correct object in the ironic trials will be higher in the dissimilar condition than in the similar condition, since a greater number of looks conveys more uncertainty in decision-making.
8. If similarity is a cue to speaker intent for irony, then the proportion of total fixation duration dedicated to the correct object will be higher in the similar condition than in the dissimilar condition.

Method

Participants

Participants were 30 five-and six-year-old children (*M* = 5.55 years, *SD* = 0.50, 15 female), 20 seven- and eight-year-old children (*M* = 7.56 years, *SD* = 0.50, 10 female), and 20 adults (*M* = 22.92 years, *SD* = 7.58, 16 female). The participants were recruited and rewarded in the same manner as the participants in Experiment 1. Two additional participants were tested but were not included in the final analyses due to problems with the video recording.

Materials

Puppet shows. Ten puppet shows, each involving two puppets and props, were used to depict stories involving the two puppets. The speaker puppet in each story was one of two puppets that the participant selected at the start of the session: either a puppet that the participant felt was similar to them or a puppet the participant felt was dissimilar from them. The stories were the same as those used in Experiment 1, except that one story from Experiment 1 (about shopping) was replaced with a different story (about a trampoline) that was meant to be more gender-neutral and appealing to all participants.

Each puppet show started with a statement that identified the puppets, "Shannon and Lisa just met. They play on a soccer team together…", and ended with an evaluative statement made by the speaker puppet. As with Experiment 1, the evaluative statement was a literal compliment for two of the stories, a literal criticism for two of the stories, and an ironic criticism for six of the stories. Participants made evaluations of speaker intent using the shark, duck and answer box as described previously. Appendix A contains sample puppet show stories.

Language measures. The PPVT-4 (Dunn & Dunn) was administered to all child and adult participants to measure receptive vocabulary as an indicator of linguistic proficiency. The CCC (Bishop, 1998) was administered to the parents of child participants as a measure of pragmatic language understanding.

Theory of mind measure. Two false belief stories were administered (in counterbalanced order) at the end of each session to the child participants only.

Demographic questionnaire. Parents were provided with a brief demographic questionnaire to elicit information about the participants' number of siblings, birth order, sibling genders, and other relevant factors. See Appendix C for the questionnaire.

Procedure

The only notable difference between the experimental procedure of Experiment 1 and Experiment 2 was that each participant had to select a similar and dissimilar puppet to act as speakers in the puppet shows in Experiment 2. The pertinent details of this selection process will be outlined here, but all other procedural details are the same as those of Experiment 1.

During the warm-up portion at the start of the participant's testing session, the

experimenter asked the participant several questions about their hobbies, pets, favorite sports and school involvement. These questions were posed in a conversational manner while engaging with the participant and their parent (or just with the participant in the case of adult participation). The answers to these questions were used, along with obvious physical characteristics, to identify five attributes the experimenter could reference during the puppet selection process. While the participants completed relevant paperwork, the experimenter excused herself from the room and made note of these characteristics on the coding sheet so they could be referenced in the puppet introductions. At this time, the experimenter also selected two puppets from a larger selection: one puppet that was the same gender as the participant and matched the participant's hair color, length and texture, and skin tone most closely, and one puppet that was the opposite gender from the participant and was as disparate in physical appearance as possible. Similar procedures have been used successfully in past studies (e.g., Reyes-Jacques & Echols, 2013).

When the testing session began, the experimenter told the participant she wanted them to meet two of the puppets, Emma and Shane, who would be in the puppet shows. The experimenter introduced each of the two puppets to the participant by providing the relevant characteristics. The experimenter always commented on 1) gender, 2) age/grade, 3) presence/absence of siblings, 4) hair color/texture, 5) favorite hobby, sport or pet. When introducing the 'similar' puppet the experimenter noted the similarity between the participant and puppet after stating each characteristic (e.g., "Emma has long blonde hair, just like you!"), and when introducing the 'dissimilar' puppet the experimenter noted the dissimilarity between the participant and puppet after stating each characteristic (e.g.,

“Shane has short brown hair; that’s not at all like you.”). The introduction of the alike puppet and the different puppet was counterbalanced across participants. After introducing both puppets, the experimenter asked the participant to choose a puppet that they felt was most like them and to choose a puppet they felt was not at all like them. The experimenter then proceeded with the 10 puppet shows, using Emma as the speaker in half of the shows, and Shane as the speaker in half. At the end of the final puppet show, the experimenter posed a manipulation question to the participant to ensure they remembered which puppet they had selected as being similar to them and dissimilar from them: “At the beginning of the puppet shows today you told me one of these puppets was like you and one of these puppets was not like you. Can you tell me again, which puppet did you think was a lot like you? And which puppet did you think was not at all like you?” All participants passed this manipulation check by identifying the same similar and dissimilar puppets at the end as they had in the beginning.

Design. The experimental design was the same as that of Experiment 1, except that no sibling or stranger relationship information was provided, and instead the speaker puppet on half of the trials was the puppet the participant stated was similar to them and on the other half of the trials it was the puppet they stated was dissimilar from them. Across participants every puppet show was presented with every condition. As with Experiment 1, 10 versions of the materials were created.

Despite efforts to counterbalance the pairing of each language type with the presentation of each story, a significant counterbalancing error occurred in the creation of the versions for this experiment, which was not discovered until after testing had concluded. This counterbalancing issue caused 7 of the 10 protocol versions to contain an

imbalance of trials that were divided amongst the similar speaker and the dissimilar speaker. In each of the 10 versions exactly 6 of the story endings were ironic criticisms, 2 were literal criticisms, and 2 were literal compliments, just as intended. In addition, exactly half of the stories in each version featured the similar speaker and the remaining half featured the dissimilar speaker, just as intended. The imbalance occurred on 7 of the protocol versions when the language endings and speaker similarity factors were not rotated through all the desired combinations correctly. For instance, version 1 contained two stories that featured Emma and ended in a literal criticism, but no stories that featured Shane with a literal criticism ending. There should have been one literal criticism story for each speaker so that participants who chose Shane as being like them and Emma as being different from them (i.e., male participants) had the Shane story contribute to the alike condition and the Emma story contribute to the different condition. Across all 10 versions the same combinations of stories and endings are represented, but the imbalance within any given participant meant that even more data were missing in this experiment.

A second coder examined 25% of the video data to ensure accuracy in measurement of the response latency and eye gaze data. The Crohnbach's alphas demonstrated good inter-rater reliability for the number of frames in each of the three response phases, with agreement from the start of the statement to the beginning of the reach being $\alpha = 1.00$, from the start of the reach until the participant touched the response object being $\alpha = 0.86$, and from touching the object to placing it in the answer box being $\alpha = 0.98$. Similarly, the alphas for eye gaze data showed high reliability between raters for the number of fixations to the shark ($\alpha = 0.98$) and to the duck ($\alpha = 0.99$), as well as

for the duration of time spent looking at the shark (α = 0.97) and the duck (α = 0.99).

Results

This results section will address data from adult participants first, followed by results from analyses of children's data. For both the adults and the children, results will be described first for the behavioral measures of accuracy, followed by explication of the response latency data, and finally an analysis of the eye gaze data. For adult participants a series of paired-samples t-tests were run to test for differences based on speaker similarity (similar, dissimilar); for child participants a series of 2 x 2 mixed-model ANOVAs were run to test for differences based on similarity (a within-subjects factor) and age group (5- and 6-year-olds, 7- and 8-year-olds; a between-subjects factor). In order to be certain that participants were attending to the speaker similarity information that was presented at the start of the testing session, a manipulation check question was posed at the end of the puppet shows. The experimenter asked the participant to identify which of the two speaker puppets had previously been selected as being similar to the participant and which puppet had previously been selected as being dissimilar from the participant. All adult and child participants responded correctly to this question so no cases were removed due to this question before proceeding with the analyses.

As with Experiment 1, only subjects analyses were conducted. However, Table 16 provides an overview of accuracy on each item for visual inspection. In this experiment, adults' accuracy on the "dessert" story was lower than might be expected, but children's accuracy on this story was strong so it was kept in the analyses.

Table 16

Participants' mean speaker intent accuracy on ironic criticism trials for each story (item)

Story	5-6 year olds	7-8 year olds	Adults
Soccer	.78	.86	.91
Nintendo	.71	.57	.78
Dessert	.60	.60	.25
Stampede	.88	1.00	.92
Trampoline	.70	.91	1.00
Waterskiing	.56	.60	.91
Hide & Seek	.57	.80	1.00
Bikes	.78	.60	1.00
Snowboarding	.91	.89	1.00
Painting	.71	.80	1.00

Note. Includes correct speaker belief trials only.

Adult participants

Speaker belief. When the speaker made a literal compliment the correct evaluation was that the speaker believed the outcome was good, whereas when the speaker made a literal criticism or an ironic criticism the correct evaluation was that the speaker believed the outcome was bad. Adults' performance on the speaker belief question is summarized in Table 17.

Table 17

Adults' mean proportion correct on the speaker belief question

Ironic Criticism		Literal Criticism		Literal Compliment	
Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
n = 20	*n* = 20	*n* = 16	*n* = 16	*n* = 17	*n* = 15
.86 (.21)	.94 (.16)	1.00 (.00)	1.00 (.00)	1.00 (.00)	1.00 (.00)

Note. Standard deviations in parentheses. For literal criticisms and ironic criticisms the correct evaluation was that the speaker believed the outcome was bad; for literal compliments the correct evaluation was that the speaker believed the outcome was good.

A paired-samples t-test for the ironic trials showed that there was no significant difference in speaker belief accuracy between the similar condition (Mean proportion correct = .86, *SD* = .21) and the dissimilar condition (*M* = .94, *SD* = .16), $t(19) = 1.75$, $p = .096$. Adults were highly accurate at determining that the ironic speaker held a belief that the outcome was a negative one, regardless of whether they perceived the speaker as being similar to them or dissimilar from them. One-sample *t*-tests were conducted to test whether adults' speaker belief accuracy was significantly different from zero, and these tests showed that speaker belief judgments were significantly greater than zero percent accuracy for both the similar condition, $t(19) = 18.06$, $p < .001$, and the dissimilar condition, $t(19) = 26.26$, $p < .001$.

Speaker intent- aggression. A participant must understand that the ironic speaker holds a counterfactual belief about the event outcome in order to accurately determine the speaker's intent. Therefore, only trials for which there were correct evaluations on the speaker belief question were included in the speaker intent analyses.

When the speaker made ironic criticisms the correct evaluation was that the speaker intended to be mean. The proportions of correct speaker intent judgments are listed in Table 18.

Table 18

Adults' mean proportions correct on the speaker intent question

Ironic Criticism		Literal Criticism		Literal Compliment	
Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
n = 20	*n* = 20	*n* = 16	*n* = 16	*n* = 17	*n* = 15
.98 (.11)	.84 (.20)	.94 (.25)	.75 (.41)	1.00 (.00)	.96 (.13)

Note. Standard deviations in parentheses. For literal criticisms and ironic criticisms the correct evaluation was that the speaker intended to be mean; for literal compliments the correct evaluation was that the speaker intended to be nice.

A paired-samples *t*-test showed there was a significant difference between adults' accuracy at determining the speaker's ironic intent between the similar condition (M = .98, SD = .11) and the dissimilar condition (M = .84, SD = .20), $t(19) = 2.70$, $p = .014$, d = 1.42. Adult participants were better able to appreciate the speaker's ironic intent when they perceived the speaker to be similar to them than when they perceived the speaker to be dissimilar from them. One-sample *t*-tests were conducted to test whether adults' speaker intent accuracy was significantly different from zero percent accuracy, and these tests showed that speaker intent judgments were significantly greater than zero for both the similar condition, $t(19) = 39.00$, $p < .001$, and the dissimilar condition, $t(19) = 18.63$, $p < .001$.

Speaker intent- humour. As in Experiment 1, only trials on which there were correct evaluations on the speaker belief question and the aggressive intent question were included in the speaker intent to be humorous intent analyses, for both adults' and children's data. When the speaker made ironic criticisms the correct evaluation was that the speaker intended to be humorous. Adults' accuracy on the humour intent scale is listed in Table 19.

Table 19

Adults' mean proportions correct on the humour intent question

Ironic Criticism		Literal Criticism		Literal Compliment	
Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
n = 20	*n* = 20	*n* = 15	*n* = 13	*n* = 17	*n* = 14
.44 (.43)	.49 (.42)	.93 (.26)	1.00 (.00)	.68 (.47)	.79 (.43)

Note. Includes correct trials only. Standard deviations in parentheses. For literal criticisms and literal compliments the correct evaluation was that the speaker intended the remark to be serious; for ironic criticisms the correct evaluation was that the speaker intended the remark to be funny.

The paired-samples *t*-test showed no significant effect of relationship, $t(19) = 0.73$, $p = .477$. Overall accuracy at detecting humorous intent was quite low for adults and there was no difference in accuracy of adults' humorous intent judgments between the similar condition ($M = .44$, $SD = .43$) and the dissimilar condition ($M = .49$, $SD = .42$).

Response latencies. The response latency for each trial was recorded beginning with the start of the experimenter's question (e.g., "Was Emma being like the duck or like

the shark?”) and ending when either the shark or the duck was placed inside the answer box. Response latencies were further subdivided into phases that represent early, middle or late-stage processing. The early stage included responding from the start of the question until the beginning of the participant’s reach for either the shark or the duck; the middle stage included responding from the beginning of the reach movement until the participant actually touched the selected object; the late stage included the time that elapsed from when the participant touched the object until the object was placed inside the answer box. The response latencies reported here were only analyzed for trials in which the participant got the speaker belief and speaker intent questions correct. Three separate *t*-tests were conducted, one for the ironic trials at each response latency phase. Table 20 shows adults mean response latencies for each phase of responding.

In the earliest phase of responding there was no significant difference in the response latency between the similar condition (M = 2158.97 ms, SD = 1063.59) and the dissimilar condition (M = 2331.55, SD = 725.05), $t(18) = 1.03$, $p = .317$. In the middle phase of processing there was no significant difference in the response latency between the similar condition (M = 756.13, SD = 295.17) and the dissimilar condition (M = 720.42, SD = 284.96), $t(18) = 0.34$, $p = .736$. Finally, during the late phase of processing there was no significant difference in the response latency between the similar condition (M = 795.98, SD = 143.12) and the dissimilar condition (M = 817.63, SD = 291.83), $t(18) = 0.28$, $p = .781$.

Table 20

Adults' mean response latencies (in milliseconds) during the early, middle, and late phases of responding

Response phase	Ironic Criticism		Literal Criticism		Literal Compliment	
	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
	n = 20	*n* = 19	*n* = 13	*n* = 13	*n* = 17	*n* = 14
Early	2183 (1041)	2332 (725)	3757 (3030)	2365 (463)	2108 (1036)	2038 (570)
Middle	748 (289)	720 (285)	610 (187)	663 (135)	768 (261)	718 (151)
Late	788 (144)	818 (292)	711 (126)	743 (159)	766 (101)	755 (152)
Total	3719 (963)	3870 (903)	5078 (2872)	3772 (670)	3641 (1170)	3511 (612)

Note. Standard deviations in parentheses. Includes correct trials only. Early response phase extends from the beginning of the statement "Was Emma…." until the beginning of the participant's reach; middle phase extends from the beginning of the reach until the participant touches the response object; late phase extends from participant's touch of the response object until the object is placed in the answer box to end the trial.

Eye gaze. Adult performance on each of the eye gaze measures is summarized in Table 21. First looks accuracy can provide insight into whether early processing is happening in a modular or parallel fashion. If, for instance, participants directed a significant proportion of their first looks towards the duck on ironic criticism trials, this could reflect a tendency to consider a literal interpretation first. As such, instead of testing for differences due to speaker similarity on the first looks accuracy measure, I instead conducted a one-way ANOVA to explore for differences in first looks due to language. The analysis of first looks accuracy was significant, $F(2, 36) = 5.55$, $p = .008$, $\eta_p^2 = .236$. Paired-samples *t*-tests, with a Bonferroni correction, showed the nature of this effect was such that participants were more accurate in directing their first fixation to the correct response object in the ironic criticism trials ($M = .98$, $SD = .07$) than in the literal criticism trials ($M = .74$, $SD = .39$), $t(18) = 2.74$, $p = .013$; in the literal compliment trials ($M = .95$, $SD = .15$) there was a trend for the first fixation to be more accurate than in the literal criticism trials, ($M = .74$, $SD = .39$), $t(18) = 2.19$, $p = .042$, but this effect did not hold up after the Bonferroni correction. Importantly, these differences show that adults were more accurate in directing their first fixation to the shark on the ironic trials, and indicate no evidence that adults considered a literal-first interpretation on ironic trials.

Table 21

Adults' first looks accuracy, numbers of fixations, and proportions of fixation duration

Eye gaze measure	Ironic Criticism		Literal Criticism		Literal Compliment	
	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
	n = 19	*n* = 18	*n* = 14	*n* = 12	*n* = 17	*n* = 12
First looks accuracy (mean proportion)	.98(.08)	.97 (.12)	.79 (.38)	.71 (.45)	1.00 (.00)	.88 (.31)
Number of fixations on correct object	1.06 (0.12)	1.09 (0.30)	1.07 (0.43)	1.08 (0.51)	1.00 (0.00)	1.08 (0.29)
Number of fixations on incorrect object	0.02 (0.08)	0.03 (0.12)	0.31 (0.52)	0.38 (0.48)	0.00 (0.00)	0.13 (0.31)

Note. Standard deviations in parentheses. Includes correct trials only.

The remaining eye gaze analyses were conducted to test for differences based on speaker similarity. There were no differences between the similar (M = 1.06, SD = 0.13) and dissimilar conditions (M = 1.04, SD = 0.21) in the number of fixations to the correct object (i.e., the shark), $t(16) = 0.31$, $p = .757$. For the fixations to the incorrect object there was no differences between the similar condition (M = 0.02, SD = 0.08) and dissimilar condition (M = 0.03, SD = 0.12), $t(16) = 0.28$, $p = .786$. Taken together, these eye gaze data indicate that adults are highly accurate at directing their gaze towards the object that is correctly chosen (the shark) when judging speaker intent for ironic criticisms. Adults' tendencies to consider the correct and incorrect response objects were not affected by the speaker's perceived similarity to the participant.

The proportion of looking time to each of the response objects, the shark and the duck, were considered for the three separate response phases. These proportions are listed in Table 22. Adults' strong tendency to direct their gaze to the correct object led to insufficient looks to the incorrect object in some conditions at each phase of processing. As such, only adults' proportions of looking time to the shark will be analyzed here.

Table 22

Adults' proportions of looking time to the correct and incorrect response objects during the early, middle and late phases of responding

Object	Ironic Criticism		Literal Criticism		Literal Compliment	
	Similar *n* = 19	Dissimilar *n* = 16	Similar *n* = 14	Dissimilar *n* = 12	Similar *n* = 17	Dissimilar *n* = 12
Early phase						
Correct	.11 (.11)	.12 (.10)	.17 (.14)	.14 (.12)	.14 (.26)	.09 (.11)
Incorrect	.01 (.01)	.00 (.00)	.05 (.09)	.06 (.10)	.00 (.00)	.04 (.07)
Middle phase						
Correct	.72 (.32)	.68 (.33)	.64 (.37)	.70 (.35)	.78 (.22)	.65 (.42)
Incorrect	.00 (.00)	.05 (.15)	.00 (.00)	.05 (.11)	.00 (.00)	.08 (.21)
Late phase						
Correct	.02 (.04)	.02 (.03)	.07 (.17)	.04 (.08)	.01 (.03)	.08 (.17)
Incorrect	.00 (.00)	.01 (.03)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)

Note. Includes correct trials only. Proportions of looking time to the correct and incorrect object within each response phase do not sum to 1.0 because there were other possible looking locations, such as the response box and the experimenter. Those alternative looking locations were not analyzed here because looks to those locations were not relevant to the present study.

In the early phase of processing there was no difference between the proportion of fixation duration on the shark for the similar condition (M = .11, SD = .11) and the dissimilar condition (M = .12, SD = .10), $t(18) = 0.22$, $p = .830$. Likewise, for the middle phase of processing there were no differences between the similar condition (M = .76, SD

= .28) and the dissimilar condition (*M* = .68, *SD* = .33) for the fixation duration to the shark, *t*(18) = 1.12, *p* = .279. Finally, in the late stage of processing there was no difference between the proportion of fixation duration on the shark for the similar condition (*M* = .02, *SD* = .04) and the dissimilar condition (*M* = .02, *SD* = .03), *t*(18) = 0.66 , *p* = .521.

Child participants

Children's behavioral and processing data were first subjected to analyses to check for gender effects. One significant gender effect was found among the measures of irony appreciation. As in Experiment 1, female participants were more accurate in their speaker belief assessments (Mean proportion correct = .66, *SD* = .30) than were male participants (*M* = .47, *SD* = .37), *t*(48) = 2.00, *p* = .051. Additionally, here there was a gender effect on the CCC pragmatic score such that female participants (*M* = 149.28, *SD* = 8.48) were rated by their parents as having better pragmatic understanding than male participants (*M* = 144.28, *SD* = 8.91), *t*(47) = 2.02, *p* = .049. There was also a gender effect on the theory of mind measure such that female participants (*M* = 8.48, *SD* = 1.42) had better theory of mind scores than male participants (*M* = 7.20, *SD* = 2.31), *t*(48) = 2.36, *p* = .022. The gender difference observed in the speaker belief accuracy may be due to the female participants' overall proficiency with interpreting counterfactual perspectives, as evidenced by their theory of mind and pragmatic language advantage. All later analyses will be collapsed across gender for two reasons: 1) assessments of speaker intent are more important as a marker for irony understanding and there were no gender differences on that measure; 2) given the low statistical power afforded to some of the following analyses, it was not wise to further subdivide the data based on gender given

that the gender effect was only observed for the measure of counterfactual belief understanding and not the measures of intent or processing.

Speaker belief. The proportions of children's correct speaker belief judgments are listed in Table 23.

Table 23

Children's mean proportions correct on the speaker belief question

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
5-6 year olds	.57 (.38)	.50 (.41)	1.00 (.00)	.95 (.22)	1.00 (.00)	1.00 (.00)
	n = 30	*n* = 30	*n* = 28	*n* = 20	*n* = 26	*n* = 24
7-8 year olds	.54 (.36)	.69 (.34)	1.00 (.00)	1.00 (.00)	1.00 (.00)	1.00 (.00)
	n = 20	*n* = 20	*n* = 16	*n* = 14	*n* = 18	*n* = 12

Note. Standard deviations in parentheses. For literal criticisms and ironic criticisms the correct evaluation was that the speaker believed the outcome was bad; for literal compliments the correct evaluation was that the speaker believed the outcome was good.

The ANOVA on speaker belief data showed that there was a significant interaction between similarity and age, $F(1, 48) = 7.02$, $p = .011$, $\eta_p^2 = .128$. Follow-up *t*-tests showed that the speaker belief accuracy between the similar condition and dissimilar conditions was no different for the younger child group, $t(29) = 1.37$, $p = .182$. For the older child group speaker belief accuracy was higher in the dissimilar condition ($M = .69$, $SD = .34$) than in the similar condition ($M = .54$, $SD = .36$), $t(19) = 2.29$, $p = .034$, but this effect was no longer significant following a Bonferroni correction. The main effects

for similarity, $F(1, 48) = 0.87$, $p = .356$, and for age, $F(1, 48) = 0.62$, $p = .434$, were not significant.

A series of one-sample *t*-tests were also conducted to test whether children's speaker belief accuracy was significantly different from zero percent accuracy. For the younger group of children these *t*-tests showed that speaker belief accuracy was significantly greater than zero for both the similar condition, $t(29) = 8.21$, $p < .001$, and the dissimilar condition, $t(29) = 6.66$, $p < .001$. Likewise, for the older group of children speaker belief accuracy was significantly greater than zero for both the similar condition, $t(19) = 6.73$, $p < .001$, and the dissimilar condition, $t(19) = 9.10$, $p < .001$.

Speaker intent- aggression. The proportions of children's correct speaker intent judgments are listed in Table 24.

Table 24

Children's mean proportions correct on the speaker intent question

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
5-6 year olds	.72 (.41)	.73 (.35)	.95 (.21)	.95 (.23)	.96 (.20)	1.00 (.00)
	$n = 24$	$n = 20$	$n = 28$	$n = 19$	$n = 26$	$n = 24$
7-8 year olds	.73 (.40)	.76 (.28)	.91 (.27)	1.00 (.00)	1.00 (.00)	.83 (.39)
	$n = 16$	$n = 18$	$n = 16$	$n = 14$	$n = 18$	$n = 12$

Note. Standard deviations in parentheses. For literal criticisms and ironic criticisms the correct evaluation was that the speaker intended to be mean; for literal compliments the correct evaluation was that the speaker intended to be nice.

These accuracy results only included trials for which the participant understood the speaker belief. Results showed that the Similarity x Age Group interaction was not significant, $F(1, 31) = 0.10$, $p = .757$; likewise, there were no main effects of similarity, $F(1, 31) = 0.53$, $p = .472$, or of age group, $F(1, 31) = 0.70$, $p = .410$. A series of one-sample *t*-tests were also conducted to test whether children's speaker intent accuracy was significantly different from zero percent accuracy. For the younger group of children these *t*-tests showed that speaker intent accuracy was significantly greater than zero for both the similar condition, $t(23) = 8.50$, $p < .001$, and the dissimilar condition, $t(19) = 9.40$, $p < .001$. Likewise, for the older group of children speaker belief accuracy was significantly greater than zero for both the similar condition, $t(15) = 7.24$, $p < .001$, and the dissimilar condition, $t(17) = 11.47$, $p < .001$.

Speaker intent- humour. The proportions of children's correct judgments of speaker intent to be humorous are listed in Table 25, and include only cases where the participant correctly identified the speaker's belief and intent.

Table 25

Children's mean proportions correct on the humorous intent question

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
5-6 year olds	.33 (.47)	.42 (.45)	.85 (.27)	.89 (.32)	.62 (.48)	.79 (.41)
	n = 19	*n* = 18	*n* = 27	*n* = 18	*n* = 25	*n* = 24
7-8 year olds	.69 (.43)	.46 (.48)	.87 (.35)	.82 (.37)	.69 (.46)	.80 (.42)
	n = 13	*n* = 17	*n* = 15	*n* = 14	*n* = 18	*n* = 10

Note. Includes correct trials only. Standard deviations in parentheses. For literal criticisms and literal compliments the correct evaluation was that the speaker intended the remark to be serious; for ironic criticisms the correct evaluation was that the speaker intended the remark to be funny.

Results showed that neither the Similarity x Age Group interaction, $F(1, 24) = 0.12$, $p = .737$, nor the similarity main effect were significant, $F(1, 24) < 0.01$, $p = .979$. However, the main effect of age was marginally significant, $F(1, 24) = 4.13$, $p = .053$. The children in the older group were more accurate at determining the humour intended with the ironic utterance ($M = .50$, $SD = .48$) than were the children in the younger group ($M = .41$, $SD = .47$).

Response latencies. Table 26 depicts the mean response latencies for each response phase (early, middle and late) for the younger and the older child groups. In the early stage of processing, there were no differences in response latency due to interactions between similarity and age group, $F(1, 23) = 0.15$, $p = .699$, or due to main effects of similarity $F(1, 23) = 0.22$, $p = .645$, or age group $F(1, 23) = 0.62$, $p = .438$. During the middle phase of processing there were no differences in response latency due

to interactions between similarity and age group, $F(1, 24) = 0.40$, $p = .536$, or due to main effects of similarity $F(1, 24) = 0.52$, $p = .479$, or age group $F(1, 24) < 0.01$, $p = .966$. In the final stage of processing, there were no differences in response latency due to main effects of similarity $F(1, 24) = 1.14$, $p = .297$, or age group $F(1, 24) = 0.04$, $p = .842$. Similarly, the interactions between similarity and age group was not significant, although the interaction did show a trend towards significance, $F(1, 24) = 3.50$, $p = .074$. Table 25 shows that the nature of this interaction appeared to be that the younger children's late-stage processing was facilitated by the presence of a dissimilar speaker.

Table 26

Children's mean response latencies (in milliseconds) during the early, middle and late phases of responding

Age Group	Ironic Criticism		Literal Criticism		Literal Compliment	
	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
			Early phase			
5-6 year olds	3314 (1997)	2395 (2275)	2050 (1873)	2766 (2783)	1476 (1829)	1911 (1399)
	n = 18	*n* = 17	*n* = 27	*n* = 15	*n* = 23	*n* = 23
7-8 year olds	2158 (1181)	2084 (1163)	1631 (1207)	2927 (2507)	1155 (1087)	1986 (1438)
	n = 13	*n* = 17	*n* = 15	*n* = 14	*n* = 17	*n* = 10
			Middle phase			
5-6 year olds	602 (177)	717 (489)	730 (393)	601 (230)	585 (153)	605 (189)
	n = 18	*n* = 17	*n* = 27	*n* = 15	*n* = 23	*n* = 23
7-8 year olds	582 (125)	653 (241)	624 (205)	618 (212)	695 (297)	518 (132)
	n = 13	*n* = 17	*n* = 15	*n* = 14	*n* = 17	*n* = 10
			Late phase			
5-6 year olds	1147 (523)	920 (398)	1124 (643)	1012 (769)	1000 (705)	1026 (975)
	n = 18	*n* = 17	*n* = 27	*n* = 15	*n* = 23	*n* = 23
7-8 year olds	1040 (482)	1029 (626)	1157 (1292)	879 (529)	800 (303)	819 (170)
	n = 13	*n* = 17	*n* = 15	*n* = 14	*n* = 17	*n* = 10
			Total response time			
5-6 year-olds	4999 (2107)	4032 (2249)	3903 (2190)	4378 (2958)	3061 (2022)	3542 (1839)
	n = 18	*n* = 17	*n* = 27	*n* = 15	*n* = 23	*n* = 23
7-8 year-olds	3779 (1342)	3766 (1468)	3412 (2030)	4424 (2579)	2651 (987)	3323 (1394)
	n = 13	*n* = 17	*n* = 15	*n* = 14	*n* = 17	*n* = 10

Note. Standard deviations in parentheses. Includes correct trials only.

Eye gaze. Eye gaze measures included the proportions of first looks directed towards the correct response object, the number of fixations on the correct object, the number of fixations on the incorrect object, and the proportion of the total fixation duration to the correct object. Only trials for which there was at least one look to the shark or the duck were included in the analyses of eye gaze measures. That is, if participants kept their gaze fixed on the experimenter and did not look to either response object, then they did not contribute any eye gaze data on that trial.

First looks. The proportion of first looks that were directed towards the appropriate response object (i.e., the shark) can be viewed in Table 27.

Table 27

Children's mean first looks accuracy

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
5-6 year olds	.82 (.38)	.94 (.17)	.96 (.20)	.93 (.27)	.92 (.26)	.95 (.22)
	n = 18	*n* = 16	*n* = 25	*n* = 14	*n* = 18	*n* = 20
7-8 year olds	.96 (.14)	.96 (.11)	1.00 (.00)	.88 (.30)	.94 (.24)	.78 (.44)
	n = 12	*n* = 16	*n* = 13	*n* = 13	*n* = 17	*n* = 9

Note. Includes correct trials only.

As in the analyses of Experiment 1, the first looks here were analyzed to determine whether eye gaze reflected consideration of the correct response object from the very earliest time point in processing: the first fixation. There were no significant differences in the proportion of first looks directed at the correct object for the language

and age interaction, $F(2, 60) = 0.96$, $p = .389$, or for the main effects of language, $F(2, 60) = 0.14$, $p = .871$, or age group $F(1, 30) = 0.80$, $p = .378$. Children were equally accurate at directing their first look towards the correct object in the ironic criticism trials as they were in directing their gaze to the correct response object in literal trials, and children's eye gaze during the first fixation did not indicate that they first considered a literal interpretation.

Number of looks. The mean numbers of fixations to the correct and incorrect response objects for correct trials are listed in Table 28.

There were no significant differences in the number of fixations to the correct object (the shark) for the Similarity x Age interaction, $F(1, 22) = 0.96$, $p = .339$, or for the main effects of similarity, $F(1, 22) = 3.17$, $p = .089$, or of age, $F(1, 22) = 0.37$, $p = .549$. Similarly, the analysis of the number of fixations to the incorrect object (the duck) demonstrated no Similarity x Age interaction, $F(1, 22) = 2.03$, $p = .168$, and no main effects of similarity, $F(1, 22) = 0.87$, $p = .361$, or of age, $F(1, 22) = 1.02$, $p = .324$.

Table 28

Children's mean numbers of fixations on the correct response object and the incorrect response object

Age Group	Ironic Criticism		Literal Criticism		Literal Compliment	
	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
			Fixations on Correct Object			
5-6 year olds	1.08 (0.19)	1.21 (0.36)	1.26 (0.41)	1.14 (0.36)	1.31 (0.46)	1.23 (0.53)
	n = 18	*n* = 16	*n* = 25	*n* = 14	*n* = 18	*n* = 20
7-8 year olds	1.08 (0.19)	1.19 (0.37)	1.19 (0.38)	1.27 (0.44)	1.32 (0.75)	1.06 (0.17)
	n = 12	*n* = 16	*n* = 13	*n* = 13	*n* = 17	*n* = 9
			Fixations on Incorrect Object			
5-6 year olds	0.27 (0.55)	0.06 (0.17)	0.12 (0.33)	0.14 (0.53)	0.08 (0.26)	0.05 (0.22)
	n = 18	*n* = 16	*n* = 25	*n* = 14	*n* = 18	*n* = 20
7-8 year olds	0.04 (0.14)	0.08 (0.26)	0.00 (0.00)	0.12 (0.30)	0.29 (0.99)	0.22 (0.44)
	n = 12	*n* = 16	*n* = 13	*n* = 13	*n* = 17	*n* = 9

Note. Standard deviations in parentheses. Includes correct trials only.

Proportion of looking time to target objects during each response phase. Analyses were conducted on the proportion of looking time to the two response objects (the shark and the duck) for each phase of responding (Table 29).

In the early phase of processing there were no significant differences in the proportions of looking time to the shark based on an interaction between similarity and age, $F(1, 22) = 1.96$, $p = .175$, or based on main effects of similarity, $F(1, 22) = 2.56$, $p = .124$, or of age group, $F(1, 22) = 0.47$, $p = .501$. Similarly, in the early phase of processing there were no significant differences in the proportion of looking time to the duck based on an interaction between similarity and age, $F(1, 22) = 3.13$, $p = .091$, or based on main effects of similarity, $F(1, 22) = 0.01$, $p = .911$, or of age group, $F(1, 22) < 0.01$, $p = .962$.

In the middle phase of processing there were no significant differences in the proportion of looking time to the shark based on an interaction between similarity and age, $F(1, 24) = 0.08$, $p = .774$, or based on main effects of similarity, $F(1, 24) = 0.22$, $p = .647$, or of age group, $F(1, 24) < 0.01$, $p = .978$. Similarly, in the middle phase of processing there were no significant differences in the proportion of looking time to the duck based on an interaction between similarity and age, $F(1, 24) = 1.83$, $p = .189$, or based on main effects of similarity, $F(1, 24) = 0.17$, $p = .684$, or of age group, $F(1, 24) = 0.17$, $p = .684$.

Table 29

Children's proportions of looking time to the correct and incorrect response object during the early, middle and late phases of responding

	Ironic Criticism		Literal Criticism		Literal Compliment	
Age Group	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
			Early phase			
5-6 year olds						
Correct	.17 (.24)	.08 (.10)	.11 (.11)	.09 (.08)	.15 (.25)	.17 (.24)
Incorrect	.04 (.06)	.01 (.01)	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.03)
	n = 18	*n* = 16	*n* = 25	*n* = 14	*n* = 18	*n* = 20
7-8 year olds						
Correct	.18 (.26)	.15 (.12)	.13 (.18)	.21 (.21)	.21 (.18)	.11 (.10)
Incorrect	.01 (.02)	.04 (.12)	.01 (.04)	.01 (.02)	.03 (.08)	.16 (.32)
	n = 12	*n* = 16	*n* = 13	*n* = 13	*n* = 17	*n* = 9
			Middle phase			
5-6 year olds						
Correct	.71 (.35)	.76 (.25)	.74 (.31)	.72 (.36)	.70 (.32)	.73 (.30)
Incorrect	.03 (.12)	.00 (.00)	.03 (.17)	.00 (.00)	.03 (.09)	.00 (.00)
	n = 18	*n* = 16	*n* = 25	*n* = 14	*n* = 18	*n* = 20
7-8 year olds						
Correct	.74 (.35)	.76 (.28)	.71 (.30)	.75 (.29)	.71 (.37)	.76 (.41)
Incorrect	.00 (.00)	.05 (.16)	.07 (.25)	.02 (.07)	.04 (.18)	.12 (.31)
	n = 12	*n* = 16	*n* = 13	*n* = 13	*n* = 17	*n* = 9
			Late phase			
5-6 year olds						
Correct	.11 (.19)	.08 (.14)	.13 (.19)	.10 (.26)	.16 (.23)	.11 (.25)
Incorrect	.01 (.01)	.00 (.00)	.03 (.11)	.00 (.00)	.00 (.00)	.00 (.00)
	n = 18	*n* = 16	*n* = 25	*n* = 14	*n* = 18	*n* = 20
7-8 year olds						
Correct	.07 (.10)	.14 (.20)	.09 (.13)	.07 (.10)	.05 (.12)	.18 (.32)
Incorrect	.00 (.00)	.00 (.00)	.02 (.06)	.00 (.00)	.02 (.06)	.01 (.02)
	n = 12	*n* = 16	*n* = 13	*n* = 13	*n* = 17	*n* = 9

Note. Standard deviations in parentheses. Includes correct trials only. Proportions of looking time to the correct and incorrect object within each response phase do not sum to 1.0 because there were other possible looking locations, such as the response box and the experimenter. Those alternative looking locations were not analyzed here because looks to those locations were not relevant to the present study.

In the late stage of processing, there was insufficient variance to compare looks to the duck based on similarity or age. For this late phase of processing there were no significant differences in the proportion of looking time to the shark based on an interaction between similarity and age, $F(1, 24) = 1.74$, $p = .200$, or based on main effects of similarity, $F(1, 24) = 0.22$, $p = .647$, or of age group, $F(1, 24) < 0.01$, $p = .978$.

Individual difference measures. Partial correlations were conducted for the child participants as a group (controlling for age), for ironic trials only (Table 30). Correlations were conducted on each of the above-reported dependent measures and the following individual difference measures: theory of mind, PPVT raw score, CCC pragmatic score, and number of siblings in family (from the demographic questionnaire).

Five correlations emerged between the behavioral measures and individual difference measures, and one additional correlation neared significance. Theory of mind was negatively related to humour accuracy, r (38) = -.34, $p = .034$, indicating that a better developed theory of mind resulted in participants rating the ironic criticisms as having a serious, rather than funny, intent. Better theory of mind enables the participant to interpret things from the target's perspective, and this may lead to an over-identification with the critical nature of the remark. Next, theory of mind also correlated with response latency in the middle phase of processing, r (37) = -.39, $p = .014$, such that a higher theory of mind score was related to faster responding. Next, theory of mind was negatively related to the number of looks to the incorrect object, r (35) = -.33, $p = .044$. Participants with higher theory of mind scores made fewer looks to the duck, perhaps because of greater certainty in their ironic interpretation.

Next, there was a positive relationship between speaker belief accuracy and the number of siblings the participant had, r (46) = .31, p = .036. Participants with more siblings were better able to correctly identify that the ironic speaker held a counterfactual belief about the outcome referenced in their remark. Finally, there were negative correlations (one significant and one nearing significance) between humour accuracy and two of the language measures. Scores on the PPVT-4 were negatively related to humour detection, r (37) = -.34, p = .035, and scores on the pragmatic subscale of the CCC were negatively related to children's' detection of humour, r (37) = -.31, p = .058, although this correlation only approached significance. That is, for both language measures greater language proficiency was associated with a tendency to judge ironic criticisms as having a serious intent rather than a humorous one.

Table 30

Partial correlations between children's performance on irony task measures and individual difference measures

Dependent measure		ToM	PPVT	CCC Pragmatic	# of siblings
Speaker Belief	*n* = 46	.22	.20	.09	.31*
Speaker Intent	*n* = 41	-.19	.10	-.12	.11
Humour Intent	*n* = 37	-.34*	-.37*	-.30	-.18
Latency-early	*n* = 36	-.23	-.08	.03	-.05
Latency- middle	*n* = 36	-.39*	-.17	-.10	-.20
Latency- late	*n* = 36	.06	.08	.12	-.01
First look correct	*n* = 34	.28	-.02	-.08	-.06
Looks to correct	*n* = 34	.04	-.09	.08	-.20
Looks to incorrect	*n* = 34	-.33*	-.08	.08	.07
Proportion time to correct	*n* = 34	.18	-.05	-.15	-.11
Shark proportion-early	*n* = 34	-.14	-.13	.01	-.03
Duck proportion- early	*n* = 34	-.01	-.03	.11	.17
Shark proportion- middle	*n* = 34	-.11	.21	-.12	-.09
Duck proportion- middle	*n* = 34	-.10	-.08	-.08	-.04
Shark proportion- late	*n* = 34	.14	.25	.24	-.02
Duck proportion- late	*n* = 34	.04	-.14	-.18	-.16

Note. Correlations reported here only include data from child participants (younger and older child groups combined) and ironic trials. Age was partialled-out of these correlations. * $p < .05$

Note that the correlation between theory of mind and the response latency during the middle phase of processing is negative in this experiment, but the same variables were correlated positively in the first experiment. In order to explore possible reasons for this difference, I conducted additional analyses that explored the relationship between theory of mind and response latency for each of the speaker cue conditions separately. All individual differences that were common across both experiments were included here. These new correlations are presented in Table 31.

Table 31

Partial correlations between children's response latency during the middle phase of processing and individual difference measures

	Experiment 1		Experiment 2	
Individual Difference Measure	Sibling *n* = 26	Stranger *n* = 21	Similar *n* = 29	Dissimilar *n* = 31
ToM	.47*	.13	-.35	-.28
PPVT	.03	-.20	<.01	-.18
CCC	-.02	.33	.17	-.08
# Siblings	.14	.27	-.09	-.15

Note. Correlations reported here only include data from child participants (younger and older child groups combined) and ironic trials. Age was partialled-out of these correlations. * $p < .05$

These additional correlations resulted in one significant correlation and one near-significant correlation. In Experiment 1 there was a positive relationship between theory of mind and response latency for the sibling condition only, $r(26) = .47$, $p = .012$. In the

condition where siblings were present, participants with a higher theory of mind score took longer to reach for the shark. In Experiment 2 there was a near-significant relationship between theory of mind and response latency for the similar condition only, $r(29) = -.35$, $p = .054$. In the similar condition, participants with a higher theory of mind score were faster to respond. Across both experiments the theory of mind and response latency correlations emerged based on the presence of a particular cue, but the cue was used in a different fashion in each experiment.

Correlations between measures in Experiment 1 and Experiment 2 combined. Using individual difference measures in conjunction with the irony task in the present set of experiments allows us to identify possible contributors to the individual differences in performance. I was seeking to identify any correlates of overall irony appreciation and processing. As such, I collapsed across the different types of speaker information in both experiments to check for correlations in both experiments combined. This approach ought to offer greater statistical power to detect effects, and may also help resolve the ambiguity present in some of the correlations that were found in each experiment separately (e.g., theory of mind being negatively correlated with response latency in one experiment and positively correlated with it in another).

When correlations were conducted on each of the individual difference measures and each of the dependent variables from the prior analyses one significant correlation emerged. Theory of mind was positively correlated with first look accuracy, $r(64) = .25$, $p = .046$, indicating that when children had a better developed theory of mind, they were more likely to make an inference that the speaker's intention was ironic and accordingly direct their first fixation to the shark. Across both experiments reported here a large

number of correlations were conducted; the observed effects should be interpreted cautiously as some of these effects could be spurious.

Discussion

The purpose of Experiment 2 was to determine a) whether perceived similarity to an ironic speaker would facilitate irony appreciation for 5- to 8-year-old children; b) how irony processing was impacted by the perceived similarity to the speaker; and c) whether individual difference measures like theory of mind and language proficiency were related to overall irony appreciation.

Adult performance on the speaker belief and speaker intent questions was quite high; adults were adept at understanding that a speaker meant to communicate ironic intent. The manipulation of the speaker similarity cue did impact adults' accuracy on the speaker intent question; adults were more accurate at identifying ironic intent when the utterance was delivered by a speaker who was perceived to be similar to the participant. It seems that perceived similarity to a speaker may in fact permit easier access to the speaker's mental states so that an ironic interpretation is more accurate when that similarity is present. Adults did not demonstrate use of the speaker similarity information in their response latencies or eye gaze behavior.

Of greater interest, of course, was whether children's irony comprehension and processing was improved when the speaker puppet was viewed as being similar to them. In the behavioral analyses there was a significant difference in 7- and 8-year-olds' understanding of the speaker's counterfactual belief across the similar and dissimilar conditions. This older age group was more accurate at determining that the speaker really thought the outcome of the situation was bad (even though they made an ironic comment

on the situation using positively-worded language) when the speaker puppet was *dissimilar* from the participant. My original prediction was that the similar condition would produce higher accuracy on this behavioral measure, but the opposite was true. No differences due to age or similarity were present for the speaker intent measure. The only other significant result among the behavioral measures was a developmental trend in humour detection, with the older group of children being marginally more accurate at detecting the humorous intent in ironic utterances than were the younger group of children.

The response latency data and the eye gaze data did not reveal any developmental differences across age groups, and only revealed one trend based on speaker similarity. Speaker dissimilarity prompted slightly faster response latencies in the final stage of processing for the younger group of children. As with the older children's performance on the speaker belief accuracy measure, it appears that children are better able to interpret irony when the speaker is *not* perceived to be similar to them. This is perhaps because children view the critical nature of the ironic utterance as being inconsistent with their own typical communicative intentions.

Notably, when the speaker similarity cue was used in interpretation, it was used differently by children than by adults. The adult participants were better able to identify the speaker's ironic intent when the speaker was similar to them. In contrast, children were better able to identify the speaker's counterfactual belief, and were somewhat faster in making responses in the late phase of processing, when the speaker was *dissimilar* to them. This contradictory pattern may be due to differences in experience with language. Children have a more limited range of experience with ironic language, and may in fact

have more experience with dissimilar speakers using irony. That is, children likely perceive parents and other adults as dissimilar to them in many ways, and these speakers may be the source of much ironic interaction for children. Adults, however, have more experience with similar (adult) speakers using irony and this experience may facilitate their ability to interpret irony more readily from similar speakers.

There are several reasons why children may not have used the speaker similarity cue more extensively in their irony processing and interpretation. As with Experiment 1, it may be the case that children relied more heavily on other cues like intonation and context incongruity, and did not need the additional cue of speaker similarity. Also, speaker similarity was established on several basic characteristics (e.g., hair color, age, hobbies) that might not be personally relevant to the participant, and therefore a strong sense of similarity may not have been established. Related to this, it may not be sufficient to be similar to someone in order to identify with their intentions; it may be necessary to share similarity of opinions or attitudes in order to have a stronger sense of identification that would provide children with insight into their mental states.

There were a few correlations of interest between dependent measures and the individual difference measures. First of all, theory of mind was negatively related to humour detection, indicating that participants with higher theory of mind scores were more inclined to interpret ironic remarks as being serious rather than funny. In addition, theory of mind was related to response latency such that participants with higher theory of mind scores were faster to respond in the middle phase of processing. When explored further, this relationship appeared to exist only for the presence of a similar speaker and not for a dissimilar speaker. Finally, theory of mind was negatively related to the number

of fixations on the incorrect object (the duck). When participants had better developed mental state reasoning ability, they were able to respond faster during the middle phase of processing and they also showed more certainty in decision making by looking to the duck less often.

Children's accuracy on the speaker belief question for ironic trials was positively correlated with the number of siblings in their home. It seems that the presence of more siblings allows children to gain experience with understanding that a speaker can say one thing while they actually believe something else. This speaker belief understanding is the first step in several milestones that make up irony appreciation. It is possible, then, that sibling presence is particularly important in the earlier years when the understanding of counterfactual belief is being mastered. These results suggest a connection between sibling presence in the home and the ability to make basic inferences about a speaker's mind. In addition, two correlations emerged between language measures and humour accuracy ratings for ironic criticisms. For both the PPVT (a measure of receptive vocabulary) and the CCC (a measure of pragmatic language understanding), there was a negative correlation with humour accuracy detection; higher language proficiency was related to the tendency to rate ironic criticisms as being serious rather than humorous. This correlation is perplexing, particularly in the case of the pragmatic language measure since a well-defined pragmatic language understanding ought to assist a participant with understanding the higher-order goals and functions of irony. Children with strong pragmatic language skills ought to be the very ones to identify potential humour in irony correctly. One possible explanation is that children overemphasized the critical nature of the remark when considering it from the target's perspective. Alternatively, it is possible

that these children do possess the ability to infer humorous intent from the remark, but simply didn't view any humour in the utterances because the scenarios didn't reflect the kind of humour children typically enjoy. The overall rates of humour accuracy detection support this possibility since humour detection was quite low for all participants in the ironic conditions.

Several modifications could improve our understanding of speaker similarity as a cue to irony moving forward. First of all, the initial goal of this study was to explore how a sense of personal identification with a speaker could facilitate irony appreciation. The manipulation used here relied upon a sense of perceived similarity between the participant and speaker, but there is no guarantee that perceived similarity is enough to ensure that the participant felt they could actually identify with the speaker. The manipulation used here was also very limited and may not reflect the depth of similarity felt between two people in real life situations. Having the participant supply information about a speaker that they truly identify with could make for a stronger manipulation. Alternatively, having participants provide a subjective rating of perceived similarity and perceived identification would allow for clearer interpretation about how a potential similarity cue is being used. Finally, establishing similarity based on shared opinions and attitudes ought to provide participants with a stronger sense of identification and may facilitate interpretation of others' mental states.

General Discussion

In the present set of experiments, two different speaker characteristics were examined to determine whether they are reliable cues to ironic intent for children. Results showed that these cues, knowledge of sibling relationship status and one's perceived

similarity to a speaker, were used in only very modest ways in determining ironic intent. One interpretation of this result is that sibling relationship information and perceived similarity are not cues that children rely upon in the process of irony interpretation. A different interpretation is that these cues may still be relevant to children's interpretations, but that the specific manipulation of these cues and the design used in the present study were not sufficiently effective to demonstrate this. In the following sections I evaluate this latter possibility, then consider implications for theories of irony processing, our understanding of irony development, and how future studies ought to proceed.

The Use of Speaker Cues

Sibling relationship information was targeted as a potential cue to irony in Experiment 1 because sibling relationships are a familiar example of close relationships for children. Previous research established that while adults capitalize upon the closeness afforded by friendships during the process of irony interpretation (Pexman & Zvaigzne, 2004) children do not (Pexman et al., 2005). Even though children do not rely on the closeness of friendships to facilitate irony understanding, it seems plausible that they are sensitive to relationship closeness in some way. It was assumed that the mere presence of a sibling relationship in the Experiment 1 stories would be enough to convey the closeness that ought to accompany sibling relationships. However, no explicit mention was made of whether the siblings in these stories were actually close in any way. It may be the case that the potential for closeness in a relationship is not enough to establish *actual* closeness. When adults use friendship information to infer ironic intent it is likely that they assume the friends in a scenario are close. The very nature of friendship implies

closeness since the relationship is a voluntary one. The same cannot be said of sibling relationships. Siblings *do* share intimacy and are likely to be close in some ways, but the involuntary nature of the relationship does not guarantee closeness. It is possible that children in the present study were sensitive to this notion, and that the inability to infer closeness from the current contexts could be one of the contributing factors in sibling relationship information not being used more widely for irony interpretation. This issue may have been compounded by the lack of physical similarity between some of the puppet pairs. The puppets' physical differences could have indicated to participants that the speaker and target were not siblings.

There is also evidence to suggest that sibling relationships, although potentially close, may not be optimal for encouraging irony interpretation. Massaro et al. (2012) found that stories involving ironic statements made by a mother were more accurately interpreted than stories involving ironic statements made by a sibling. The authors suggested that this effect emerged because participants recognized that adults demonstrate greater communicative competence than children. In particular, children likely have more experience hearing forms of indirect speech from an adult, like their mother, than from another child. This experience then facilitates irony detection in other contexts, like the scenarios the children were presented with in the Massaro et al. study. If Massaro et al.'s reasoning is correct, it might be the case that children only rely upon closeness (for instance, with a mother or sibling) when there are other supporting characteristics, such as communication skill.

Similar to the problems noted above, regarding the assumption that sibling relationships would be viewed implicitly as being close relationships, there were some

limitations in Experiment 2 regarding the way the variable of interest was defined. My original intention was to manipulate the degree to which identifying with a speaker assists someone with irony interpretation. To do this, I manipulated the perceived similarity to a speaker. However, perceived similarity and identification are not the same thing; it is possible to be quite similar to someone and recognize those similarities but not identify with the person. In the case of children in the Reyes-Jacques and Echols (2013) study, children used perceived similarity with a speaker when making assessments of speaker reliability. In that context, children used similarity to identify which speaker had access to more knowledge of a fact. However, in the present study, children have to judge more than a mere fact. They need to move beyond fact to make an inference about underlying speaker attitudes and intentions. It is possible then, that similarity is recruited for tasks that are relatively straightforward, but perhaps children cannot rely on similarity as well when the task is much harder (i.e., they need to infer underlying intentions). The ability to use speaker similarity in irony interpretation may be limited to more authentic communication environments where the need to use speaker similarity to infer attitudes and intentions is more pressing and relevant.

My goal with the present study was to induce a sense of identification, as I felt that would give insight into the mind of a speaker. If a strong sense of similarity were established, it might still be possible to use similarity as a proxy for speaker identification. In this experiment I established similarity based on basic characteristics, such as the speaker being in the same grade as the participant or having the same hair color as the participant. These are similarities of physical appearance or fact; they are not about underlying attitudes or beliefs. It may be the case that when two people share a

similarity of beliefs they are better able to interpret ironic language because the similarities between the conversational partners exist at the level of internal mental states. To understand if similarity is really a useful cue to ironic intent, manipulating similarity in this way might be more fruitful than the approach taken in the present study.

Finally, the use of speaker cues was modest in both studies but was also asymmetrical. The similarity cue was recruited across more measures in Experiment 2 than the speaker cue was in Experiment 1. This difference may relate to the predictability of the speaker in some way. Recall that in the first experiment the sibling or stranger that made the ironic utterance differed in every story; On the 6 ironic trials the participant saw 3 *different* sibling speakers and 3 *different* stranger speakers. In contrast, in the second experiment the similar or dissimilar puppet that made the ironic utterance was the same across each story; the participant saw 3 stories involving the *same* similar speaker and 3 stories involving the *same* dissimilar speaker. An unintended byproduct of this design was that participants in Experiment 2 had the opportunity to gain some familiarity with each of those speakers and could anticipate some predictability in their response styles. These participants could learn, for instance, that each of those speakers was likely to use irony with some consistency. I believe this familiarity with each speaker could have lessened task demands that in turn permitted better use of the cue. In addition, the repeated exposure to the same speaker likely reflects naturalistic conversations more than the constant change in speaker seen in Experiment 1, and may be a better reflection of how speaker cues are used in authentic communication.

Processing Findings

A second goal of the present study was to understand how children's processing of ironic language could inform us about relevant theories. There are two theoretical camps that are relevant to irony processing: modular (or literal-first) accounts (Giora, 1997), and interactive accounts (Gibbs, 1986; Katz, 2005; Pexman, 2008). These theories make very different claims about the earliest stages of interpretation. In the present study, the very first look a participant directs towards an object is the first indication we have about the meaning they are considering in the earliest measured moments. Across both experiments, children's tendency to direct their first look towards the correct object was equal for literal criticisms and ironic criticisms. The fact that children looked to the shark first in the ironic criticism condition, just as they did in the literal criticism condition, indicates that they were giving early consideration to the ironic meaning. Had they been making a literal interpretation first, their first look ought to have been directed to the duck in the ironic criticism condition. The early bias to look at the shark supports the interactive models of irony interpretation. Interestingly, adult participants in Experiment 2 did demonstrate a difference in first-looks accuracy, but this difference showed that they looked to the duck (incorrect response object) first on a greater proportion of trials for the *literal* criticism than for the ironic criticism. Adults, like children, looked to the shark on their first glance on ironic trials, which demonstrates an early bias towards ironic interpretation. This pattern of results mimics those found by Climie and Pexman (2008), and indicates that, at least based on this metric, there is no evidence that children are bound to literal-first, two-stage processing.

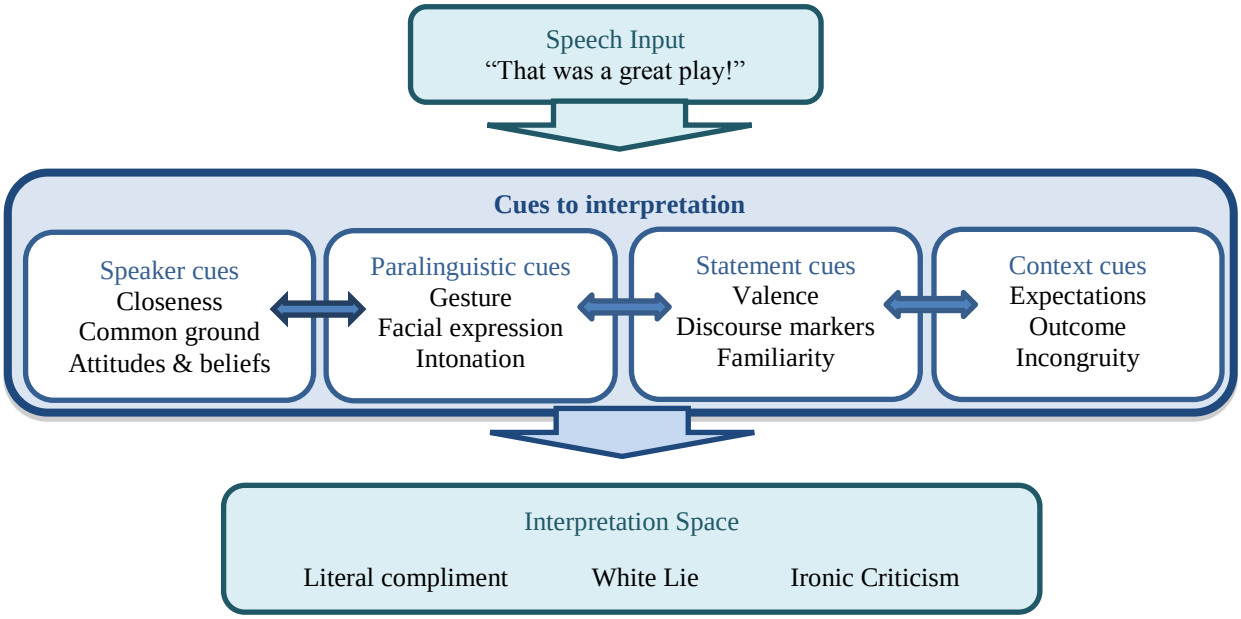
There is an important caveat to the above interpretation. The early stage of processing measured in the present study does not capture the very earliest time at which a participant could begin to make an interpretation. Recall that processing was measured starting with a question the experimenter posed after the final utterance. In the soccer scenario when Shannon says “That was a great play,” processing was measured from the start of the subsequent question: “Was Shannon being like the shark or like the duck?” The participant was given all the information they need in order to make a decision about literal or ironic interpretation once they heard the word “great” in the final statement. Seconds passed between that time and the point at which they heard the experimenter’s question. This was particularly true on early trials where participants tended to wait for (and need) the experimenter’s prompt, but was less the case on later trials when many participants began to anticipate the prompt and respond earlier. As such, participants had some time to process the remark before directing their gaze to either of the response objects. Thus, the early moments of processing in the present study are only early in relation to the middle and later stages of processing, and do not reflect the absolute earliest time that interpretation was possible. Despite this limitation, children still coordinated cues about context, intonation, and speaker information rapidly to make an inference about ironic intent. The early bias in eye gaze towards the shark indicates that an ironic interpretation was being considered *relatively* early in the process, and this evidence suggests that interactive accounts are likely suitable for interpreting developmental findings. In addition, the trend for both adults and children in Experiment 1 to have faster latencies in the first response phase for ironic statements made between siblings rather than those made between strangers indicates that consideration of this

relationship cue was occurring early in processing. This evidence of early consideration of the cue is consistent with an interactive account of processing.

Nonetheless, future research should involve a revised procedure that captures children's eye gaze to the response objects as the ironic or literal statement unfolds. This could be accomplished by training children to make an evaluation of the speaker puppet without relying on a prompt from the experimenter. For instance, an experimenter could emphasize the need for children to make an evaluation quickly, and could encourage children to make their choice about the speaker's comments being nice or mean as soon as they can as the language unfolds. In addition, if the ironic language was more consistent across trials (in length and wording) it would be possible to compare processing during the statements themselves, as children could in principle begin to derive an interpretation during (and not at the end of) the statement. Furthermore, the use of a video camera to capture eye gaze and response latency is adequate for crude measurement of processing, but has limitations. Eye tracking systems that provide more fine-grained measurement would permit more precise time-course analysis and for more detailed examination of which visual cues the participant is considering at each moment in processing.

Several other steps are needed in order to develop interactive processing accounts further. Interactive processing accounts can be conceptualized as having three broad phases of utterance understanding: the input (i.e., the utterance), the cues to interpretation, and the interpretation space (e.g., literal or ironic). Figure 3 is adapted from Pexman (2008) and shows the relationship of these three phases.

Figure 3. Model of Interactive Processing.



Systematic examination of elements within each of these three aspects of utterance understanding is required for theory development. For the first aspect, the speech input, this means conducting studies that examine different forms of nonliteral language, such as lexicalized irony, novel irony, hyperbole, rhetorical question and any other neighboring speech form that shares features with counterfactual irony. For the second aspect, the cues to interpretation, a systematic approach to studying cues is required. The present study examined just two cues, and those two cues related specifically to the speaker. Several other cues exist, such as paralinguistic features, context, expectations of a situation's outcome, politeness and social norms, emotional valence, and many others. Several of these elements have been examined developmentally: forms of irony (e.g., Hancock et al., 2000; Harris & Pexman, 2003), intonation (Capelli et al., 1990; Winner, 1988), personality trait information (Pexman et al., 2006), and relationship (Massaro et al., 2012; Pexman & Zvaigzne, 2004). However,

we know very little about how the cues are considered together. Interactive processing accounts rely on the assumption that several cues can be considered at once (Pexman, 2008) but the specific nature of how those cues interact during processing is underspecified. For the third aspect, the interpretation space, closer consideration of alternate interpretations is required. For instance, a listener may erroneously interpret a remark intended as ironic as instead being a white lie or as being literal. These instances of failed irony may be informative about the use of cues in irony interpretation. For example, when an ironic criticism is presented, if children make the decision to select the duck (inaccurately) quickly, and without directing their gaze towards the shark at all, then we could conclude they have no sense that irony was intended but instead made a white lie interpretation. If, however, they selected the duck but deliberated between the two response objects before doing so, we could conclude they were weighing the alternatives in the decision making process. This response pattern could be informative in determining which circumstances lead to certainty in the decision making and which circumstances lead to uncertainty. In order to advance interactive processing accounts, each of the above suggestions needs to be explored within contexts that permit examination of real-time processing.

Individual Differences in Irony Appreciation

A third goal of the present study was to identify potential links between irony appreciation and other measures of development, such as theory of mind and language proficiency. The results showed limited correlations between these measures in each experiment. One finding of note was the contradictory pattern between response latency and theory of mind, such that RT in the middle phase of processing was slower for

participants with strong theory of mind in Experiment 1, but processing in that same phase was faster for participants with strong theory of mind in Experiment 2. This result may be due to the notion of speaker predictability noted above. In Experiment 1 participants with higher theory of mind may show slower processing because they are using their mental state reasoning abilities to consider both perspectives in the speaker-target relationship and have to work to do this anew with each new pair of characters. When this correlation was explored further the relationship between theory of mind and response latency was found only for the sibling condition and not for the stranger condition. It seems plausible that the participants with stronger theory of mind skills may be inclined to compare each sibling pairing in the stories to their own sibling relationship as a way of understanding the likelihood of a literal versus ironic interpretation. This comparison process would take time and would result in slower response latencies.

In Experiment 2, however, the participants can more readily make judgments of intent because they have repeated exposure to the two speakers (one similar and one dissimilar) and can therefore make interpretations about the speaker's intentions more readily. When the theory of mind and response latency correlation was explored further for Experiment 2, the relationship was only found to approach significance for the similar condition and not for the dissimilar condition. This means that participants with higher theory of mind showed a facilitated processing when the speaker was like them. Unlike in Experiment 1, the participants here could readily infer perspective because of the perceived similarity to the speaker. Correlational analyses were also conducted for the data from Experiments 1 and 2 combined, using the variables that were common to both experiments. The result of this analysis showed that theory of mind was related to

accuracy in first looks; children with sharper mental state reasoning abilities were able to infer an ironic interpretation of the remark from early in processing and directed their first look to the shark as a reflection of this inference.

No other correlations with irony appreciation emerged across the two experiments when the data were combined, despite the fact that previous research suggests children's theory of mind and language proficiency are related to irony appreciation. Filippova and Astington (2008) found that both theory of mind and vocabulary were reliable predictors of children's irony understanding. Contrary to Filippova and Astington, however, there was no evidence from the present study that receptive vocabulary was helpful in irony understanding. Moreover, the role of theory of mind wasn't prominent in the present study. It is possible that the differences in irony tasks, between the present study and that of Filippova and Astington, are partly responsible for this difference. In the Filippova and Astington study ironic intent was measured by asking the participants what the story characters meant and why the characters would speak that way, rather than asking if their intent was to be nice or mean (as in the present study). The Filippova and Astington question framing seems to tap into mental state reasoning more than the question framing in the present study, and mental state reasoning is also what theory of mind questions assess. It seems likely, then, that Filippova and Astington created irony comprehension questions that are more closely aligned with theory of mind questions, and it is possible their results demonstrating the link between irony appreciation and theory of mind are due in part to this overlap. That is not to say a connection does not exist between theory of mind and ironic language interpretation. It is likely, however, that the link is

underspecified at this time and more attention needs to be given to how theory of mind relates to the different aspects of irony appreciation.

A gender difference was observed in both experiments that demonstrated female participants were marginally more accurate in making speaker belief judgments than were male participants. This effect was not observed in the judgments of the speaker's intent to be aggressive or to be humorous, and was not observed in any of the processing measures. However, in Experiment 2 female participants also outperformed male participants on measures of pragmatic language understanding and theory of mind. Past research has demonstrated a female advantage on theory of mind, particularly where emotion aspects of theory of mind are concerned (Bosaki & Astington, 1999). Despite this, the gender advantage observed in the theory of mind literature is believed to be quite modest (Charman, Ruffman, & Clements, 2002). Early sibling interaction has been credited with providing girls with better emotion understanding than boys (Brown & Dunn, 1996) and this may be relevant in particular when considering ironic language interpretation in the context of sibling relationships. The present study did not control for gender composition in the participants' families, but tighter control of this aspect in future studies may allow us to better understand the role of mental state talk amongst siblings in irony understanding.

Developmental Findings

There was a surprising lack of developmental findings in the present set of experiments. The 5- to 8-year-old age range was selected to provide a timeframe that spans the beginning of irony comprehension (Ackerman, 1983; de Groot et al., 1995; Dews et al., 1996; Winner & Leekam, 1991) and the subsequent years where

improvements in understanding are normally observed (Filippova & Astington, 2008; Hancock et al., 2000; Harris & Pexman, 2003). It was my expectation that developmental differences in the implicit understanding that could be reflected in processing measures may even be more pronounced than developmental differences in the explicit understanding typically seen during this age range. Despite this expectation, age effects were very limited. It is possible that statistical power limitations prevented the observance of effects that may be small or even moderate. Alternatively, it is possible that irony appreciation is very slow to develop and that age effects are not pronounced for that reason. Indeed, past research indicates that irony appreciation continues to unfold across middle childhood and into adolescence and adulthood (Capelli et al., 1990, Demorest et al., 1983; Dews et al., 1996). The use of more sensitive measures in future studies may permit clearer understanding of unfolding irony appreciation in later childhood.

Future Directions

Although statistical power seems to be a real limitation of the present study, it is my contention that a bigger limitation is the use of simple stories that may lack contextual richness. Although the stories used here were intended to be simple so that participants could readily complete the tasks, it could be that the simplicity of the stories and the impoverished communicative setting eliminated the need for participants to rely on some of the cues provided. In more authentic communications, knowledge about a speaker's sibling relationship or about one's own similarity to a speaker might be more useful in determining the appropriate interpretation of an utterance.

Several studies indicate that emergent irony understanding can be observed in naturalistic settings at a younger age than is often seen in laboratory settings. Pexman et al. (2009) established that children use verbal irony in the context of a cooperative task as early as age 5, and they demonstrated that some children use gestural irony even earlier, at age 4. Similarly, Recchia et al. (2010) found that children as young as age 4.5 used different forms of irony in the context of family conversations. These children used sarcasm less frequently than other nonliteral forms, like hyperbole and rhetorical question, but demonstrated a tendency to use irony appropriately in a naturalistic setting nonetheless. Children's productions can sometimes reflect adult-like competency in naturalistic settings as well; Colston (2007) established that young children's spontaneous hyperbole productions mimic those of adults in many ways, including their use of topics discussed and the way in which hyperbolic comparisons were made. Taken together, these results indicate that there is a richness in communicative context that can be exploited for the study of ironic language, and more importantly, that children's abilities to communicate with irony may emerge at a much younger age than is depicted in the studies that take place in experimental laboratory settings. In particular, it may be the relationships with experienced speakers that facilitate the appropriate use of irony in real life, as indicated with Massaro et al.'s finding that children's success with irony is better in the presence of a mother than the presence of a sibling.

Despite limitations, the results of the present study demonstrate that children do attend to speaker cues in a subtle way, and can use those cues to assist with interpretation. The results also provide support for the notion that children's irony processing is best characterized by an interactive model. In addition, the results suggest that at least one

other important developmental achievement, theory of mind, is related to irony appreciation in some ways. Certainly, there is much left to learn about development of this complex aspect of social communication. Studying this topic within the context of processing is critically important to understand how conversations are successfully co-constructed through the coordination of interpretation cues. The challenge is to create tasks that allow access to children's processing and at the same time provide rich and natural contextual information. An understanding of how these cues are combined during utterance resolution will permit more thorough theory development.

References

Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology, 35*, 487-508.

Akimoto, Y., Miyazawa, S., & Muramoto, T. (2012). Comprehension processes of verbal irony: The effects of salience, egocentric context, and allocentric theory of mind. *Metaphor and Symbol, 27,* 217-242.

Ames, D. L., Jenkins, A. C., Banaji, M . R., & Mitchell, J. P. (2008). Taking another person's perspective increases self-referential neural processing. *Psychological Science, 19,* 642-644.

Astington, J. W. (1998). Theory of mind, Humpty Dumpty, and the icebox. *Human Development, 41,* 30-39.

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false belief understanding and their ability to reason about evidence. *New Ideas in Psychology, 20,* 131-144.

Bishop, D. V. M. (1998). Development of the Children's Communication Checklist (CCC): A method for assessing qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry, 39,* 879-891.

Bosaki, S., & Astington, J. W. (1999). Theory of mind in preadolescence: Relations between social understanding and social competence. *Social Development, 8,* 237-255.

Brown, J. R., & Dunn, J. (1996). Continuities in emotion understanding from three to six years. *Child Development, 67,* 789-802.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language use*. Cambridge: Cambridge University Press.

Bryant, G. A. (2011). Verbal irony in the wild. *Pragmatics and Cognition, 19,* 291-309.

Bryant, G. A. & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech, 48*, 257-277.

Capelli, C. A., Nakagawa, N., & Madden, C. M. (1990). How children understand sarcasm: The role of context and intonation. *Child Development, 61,* 1824-1841.

Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development? *Social Development, 11*, 1-10.

Clark, H. H. (1992). *Arenas of language use.* Chicago, IL: University of Chicago Press.

Climie, E. A., & Pexman, P. M. (2008). Eye gaze provides a window on children's understanding of verbal irony. *Journal of Cognition and Development, 9,* 257-285.

Colston, H. L. (1997). Salting a wound or sugaring a pill: The pragmatic functions of ironic criticism. *Discourse Processes, 23,* 25-45.

Colston, H. L. (2002). Contrast and assimilation in verbal irony. *Journal of Pragmatics, 34*, 111-142.

Colston, H. L. (2007). What figurative language development reveals about the mind. In A. C. Schalley & D. Khlentzos (Eds.), *Mental States, Volume 2: Language and Cognitive Structure* (pp. 191-212). Amsterdam: John Benjamins Publishing.

Colston, H. L., & Keller, S. B. (1998). You'll never believe this: Irony and hyperbole in expressing surprise. *Journal of Psycholinguistic Research, 27*, 499-513.

Colston, H. L., & O'Brien, J. E. (2000). Contrast of kind versus contrast of magnitude: The pragmatic accomplishments of irony and hyperbole. *Discourse Processes, 30,* 179-199.

Creusere, M. A. (2000). A developmental test of theoretical perspectives on the understanding of verbal irony: Children's recognition of allusion and pragmatic insincerity. *Metaphor and Symbol, 15,* 29-45.

Cutting, A. L., & Dunn, J. (2006). Conversations with siblings and with friends: Links between relationship quality and social understanding. *British Journal of Developmental Psychology, 24,* 73-87.

de Groot, A., Kaplan, J., Rosenblatt, E., Dews, S., & Winner, E. (1995). Understanding versus discriminating nonliteral utterances: Evidence for dissociation. *Metaphor & Symbolic Activity, 10,* 255-273.

DeHart, G. B. (1999). Conflict and averted conflict in preschoolers' interactions with siblings and friends. In W. A. Collins, & B. Laursen (Eds.), *Relationships as developmental contexts: Festschrift in honor of Willard W. Hartup Minnesota symposia on child psychology, Vol. 30,* (pp. 281-303). Hillsdale, NJ: Erlbaum.

Demorest, A., Silberstein, L., Garnder, H., & Winner, E. (1983). Telling it as it isn't: Children's understanding of figurative language. *British Journal of Developmental Psychology, 1,* 121-134.

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

Dews, S., & Winner, E. (1999). Obligatory processing of literal and nonliteral meanings in verbal irony. *Journal of Pragmatics, 31,* 1579-1599.

Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., McGovern, A., Qualter, A., & Smarsh, B. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Development, 67,* 3071-3085.

Dress, M. L., Kreuz, R. J., Link, K. E., & Caucci, G. M. (2008). Regional variation in the use of sarcasm. *Journal of Language and Social Psychology, 27,* 71-85.

Dunn, J., & Dale, N. (1984). I a daddy: 2-year-olds' collaboration in joint pretend with sibling and with mother. In I. Bretherton (Ed.), *Symbolic play: The development of social understanding* (pp. 131-158). San Diego, CA: Academic Press.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test: Examiner's manual* (4th ed.). Bloomington, MN: Pearson Education Ltd.

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

Furman, W. & Burhmester, D. (1985). Children's perceptions of the qualities of sibling relationships. *Child Development, 56,* 448-461.

Gibbs, R. W. Jr. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General, 115,* 3-15.

Gibbs, R. W. Jr. (2000). Irony in talk among friends. *Metaphor & Symbol, 15,* 5-27.

Gibbs, R. W. Jr., & Colston, H. L. (2012). *Interpreting figurative meaning.* New York, NY: Cambridge University Press.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics, 7,* 183-206.

Giora, R., & Fein, O. (1999). Irony: Context and salience. *Metaphor and Symbol, 14,* 241-257.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3). New York, NY: Academic Press.

Hancock, J. T., Dunham, P. J., & Purdy, K. (2000). Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development, 1,* 227-248.

Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal irony. *Discourse Processes, 36,* 147-165.

Ivanko, S. L., & Pexman, P. M. (2003). Context incongruity and irony processing. *Discourse Processes, 35*, 241-279.

Jorgensen, J. (1996). The functions of sarcastic irony in speech. *Journal of Pragmatics, 26*, 613-634.

Katz, A. N. (2005). Discourse and sociocultural factors in understanding nonliteral language. In H. Colston & A. N. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (pp. 183-207). Mahwah, NJ: Erlbaum.

Katz, A. N., & Pexman, P. M. (1997). Interpreting figurative statements: Speaker occupation can change metaphor to irony. *Metaphor and Symbol, 12*, 19-41.

Kitzman, K. M., Cohen, R., & Lockwood, R. L. (2002). Are only children missing out? Comparison of the peer-related social competence of only children and siblings. *Journal of Social and Personal Relationships, 19*, 299-316.

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General, 118,* 374-386.

Kreuz, R. J., Long, D. L., & Church, M. B. (1991). On being ironic: Pragmatic and mnemonic implications. *Metaphor and Symbolic Activity, 6*, 149-162.

Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of the pie: the allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General, 124,* 3-21.

Lee, C. J., & Katz, A. N. (1998). The differential role of ridicule in sarcasm and irony. *Metaphor and Symbol, 13,* 1-15.

Massaro, D., Valle, A., & Marchetti, A. (2012). Irony and second-order false belief in children: What changes when mothers rather than sibling speak? *European Journal of Developmental Psychology, 9,* 1-17.

McHale, S., & Crouter, A. C. (1996). The family contexts of children's sibling relationships. In G. H. Brody (Ed.), *Sibling relationships: Their causes and consequences* (pp. 173-195). Norwood, NJ: Aplex Publishing Company.

Miller, S. A. (2012). *Theory of Mind: Beyond the Preschool Years.* New York, NY: Taylor and Francis.

O'Neill, D. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development, 67*, 659-677.

Perner J., & Leekham, S. (1986). Belief and quantity: Three-year-olds' adaptation to listener's knowledge. *Journal of Child Language, 13*, 305-315.

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. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science, 17*, 286-290.

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

Pexman, P. M., Glenwright, M., Hala, S., Ivanko, S., & Jungen, S. (2006). Children's use of trait information in understanding verbal irony. *Metaphor and Symbol, 21,* 39-60.

Pexman, P. M., Glenwright, M., Krol, A., & James, T. (2005). An acquired taste: Children's perceptions of humour and teasing in verbal irony. *Discourse Processes, 40,* 259-288.

Pexman, P. M., & Olineck, K. M. (2002a). Does sarcasm always sting? Investigating the impact of ironic insults and ironic compliments. *Discourse Processes, 33,* 199-217.

Pexman, P. M., & Olineck, K. M. (2002b). Understanding irony: How do stereotypes cue speaker intent? *Journal of Language and Social Psychology, 21,* 245-274.

Pexman, P. M., Whalen, J. M., & Green, J. J. (2010). Understanding verbal irony: Clues from interpretation of direct and indirect ironic remarks. *Discourse Processes, 47,* 237-261.

Pexman, P. M., Zdrazilova, L., McConnachie, D., Deater-Deckard, K., & Petrill, S. A. (2009). "That was smooth, Mom": Children's production of verbal and gestural irony. *Metaphor and Symbol, 24,* 237-248.

Pexman, P. M., & Zvaigzne, M. T. (2004). Does irony go better with friends? *Metaphor and Symbol, 19,* 143-163.

Recchia, H. E., Howe, N., Ross, H. S., & Alexander, S. (2010). Children's understanding and production of verbal irony in family conversations. *British Journal of Developmental Psychology, 28,* 255-274.

Reyes-Jacques, B., & Echols, C. H. (2013). Developmental differences in the relative weighing of informants' social attributes. *Developmental Psychology, 49*, 602-613.

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

Shapiro, J. P., Baumeister, R. F., & Kessler, J. W. (1991). A three-component model of children's teasing: Aggression, humour, and ambiguity. *Journal of Social and Clinical Psychology, 10*, 459-472.

Stauffacher, K., & DeHart, G. B. (2006). Crossing social contexts: Relational aggression between siblings and friends during early and middle childhood. *Applied Developmental Psychology, 27*, 228-240.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30,* 395-402.

Tanenhaus, M. K., & Trueswell, J. C. (2005). Eye movements as a tool for bridging the language-as-product and language-as-action traditions. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to Studying World-Situated Language Use* (pp. 3-37). Cambridge, MA: MIT Press.

Thagard, P. (2000). *Coherence in thought and action.* Cambridge, MA: Bradford Books.

Toplack, M., & Katz, A. N. (2000). On the uses of sarcastic irony. *Journal of Pragmatics, 32*, 1467-1488.

Trueswell, J. C., & Gleitman, L. (2004). Children's eye movements during listening: Developmental evidence for a constraint-based theory of sentence processing. In J. M. Henderson and F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 319-346). New York: Psychology Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72,* 655-684.

Whalen, J. M., & Pexman, P. M. (2010). How do children respond to verbal irony in face-to-face communication? The development of mode adoption across middle childhood. *Discourse Processes, 47,* 363-387.

Winner, E. (1988). *The point of words: Children's understanding of metaphor and irony*. Cambridge, MA: Harvard University Press.

Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology, 9,* 257-270.

Appendix A

Sample Stories from Experiment 1 and Experiment 2

Experiment 1

Soccer story:

Shannon and Lisa (are sisters/just met). Shannon and Lisa play on a soccer team together. It is the last few minutes of a game. Lisa kicks the ball, (scoring a goal in the last minute of the game/missing the net entirely). Shannon says:

Literal Compliment: That was a great play.
Literal Criticism: That was a terrible play.
Ironic Criticism: That was a great play.

Nintendo story:

Ethan and Mark (are brothers/just met). Ethan and Mark decide to play Nintendo one Saturday. It's Mark's turn first. Mark (moves through the first level very quickly, scoring lots of points/doesn't get past the first level of the game). Ethan says:

Literal Compliment: You are so good at this.
Literal Criticism: You are so bad at this.
Ironic Criticism: You are so good at this.

Experiment 2

Trampoline story:

Emma and Ben just met. Emma and Ben are playing on the trampoline at Ben's house. They are practicing their trampoline tricks together. Ben (does a perfect flip and lands on his feet/tries to do a flip but falls off the trampoline). Emma says:

Literal Compliment: Wow, great landing.
Literal Criticism: Wow, terrible landing.
Ironic Criticism: Wow, great landing.

Painting story:

Shane and Ava just met when Shane decided to join an art class Ava is in. One day in art class, the students are told to paint a picture of a rose. The flower Ava paints (is beautiful and looks just like a rose/ is ugly and doesn't look like a rose at all). Shane says:

Literal Compliment: You're a great artist.
Literal Criticism: You're a terrible artist.
Ironic Criticism: You're a great artist.

Appendix B

Advanced Theory of Mind Task- False Belief Stories

Chocolate bar story

This is a story about Mike and his sister Karen. Mike and Karen are both in the kitchen looking at the big chocolate bar that they were given to share. Karen eats a little bit of the chocolate bar, and it tastes really good! After she has a few bites, Karen puts the rest of the chocolate bar on the kitchen table. Then, Karen goes outside. Mike wants to make sure that he gets some chocolate too, so he hides the rest of the chocolate bar in the kitchen cupboard. Karen comes back into the kitchen and sees that the chocolate bar is gone. Karen says, "Hey Mike, where's the rest of the chocolate bar?" Remember, Mike wants to have some chocolate for himself. So he says to Karen, "I put the rest of the chocolate bar upstairs in my room." Karen says, "OK, I'll have some later." Then, Karen leaves the kitchen.

Control questions:
Where does Karen think the chocolate bar is?
Where is the chocolate bar really?

Now, Mike decides that he wants to have some of the chocolate bar. Mike goes over to the cupboard to get the chocolate bar out. Just then, Karen walks by the kitchen and she sees Mike getting the chocolate bar out of the cupboard. Karen says to herself, "Oh, Mike did not put the chocolate in his bedroom, he really put it in the kitchen cupboard." Mike does not see Karen peeking in the kitchen door.

First-order false-belief question:
Does Mike know that Karen saw him getting the chocolate from the cupboard?

Later on, Mike and Karen are both in the living room. Karen says, "I'm going to have a bit of the chocolate bar now." So Karen goes to get the chocolate bar.

Second-order false-belief question:
Where does Mike think Karen will look for the chocolate bar?

Justification question:
Why does Mike think this?

Ice cream story

This is a story about John and his friend Mary. John and Mary are playing together in the park because it is a very nice day and the sun is shining. The ice cream man is coming. Mary is sad because she really wants to buy an ice cream cone. But she left her money at home! The ice cream man says to Mary, "Don't be sad, you can go home and get money. I will be here in the park all day." So Mary goes home to get money to buy an ice cream cone. John stays in the park and plays. John sees the ice cream truck start to drive away. Remember, the ice cream man told Mary that he would be at the park all day. So John asks the ice cream man, "Hey, where are you going?" The ice cream man says, "I'm going to the church to sell ice cream. I think I can sell more ice cream at the church." So the ice cream man drives to the church to sell ice cream there.

Control questions:
Where does Mary think the ice cream truck is?
Where is the ice cream truck really?

Now, John leaves the park. Mary is at her house getting money for ice cream. Mary walks outside her house and sees the ice cream man go by. He tells Mary that he is going to the church to sell ice cream. Mary says to herself, "Well, I'm glad I know that. Now I have money so I can follow the ice cream man to the church to buy an ice cream cone." John does not see Mary talk to the ice cream man.

First-order false-belief question:
Does John know that Mary saw the ice cream man going to the church?

Later on, John goes over to Mary's house to play with her. There is a note on the door that says that Mary went to buy an ice cream cone. So John goes to find Mary.

Second-order false-belief question:
Where does John think Mary went to buy an ice cream cone?

Justification question:
Why does John think this?

Appendix C

Demographic Questionnaire

It would be helpful if you could provide some additional information about your child who is participating in today's study. Providing this information is voluntary and will be used only for the purposes of this study.

1. How many children do you have?

   _____

2. How many children currently reside in your household altogether?

   _____

3. What are the ages of each child?

   _____

4. What is the participant's birth order? (E.g., only child, first born, last born, etc.)

   _____

5. For each child in your household, please note their gender (sex) in the order they were born:

   _____