university of groningen / faculty of science and engineering

HUMAN-MACHINE COMMUNICATION
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

MASTER'S THESIS
AUGUST, 2019

Development of Irony Understanding in Typically Developing Children

Liese P. Schmidt

Internal Supervisor and Reviewer:
Dr. J. K. Spenader
Artificial Intelligence - UNIVERSITY OF GRONINGEN
First External Supervisor and Reviewer:
Prof. dr. P. Hendriks
Linguistics - UNIVERSITY OF GRONINGEN
Second External Supervisor and Reviewer:
I. Scholten
Linguistics - UNIVERSITY OF GRONINGEN

Abstract

Verbal irony is a form of non-literal language. The intended message of an ironic utterance is different from its literal meaning. The most common form of irony is ironic criticism; a positive statement is meant to be taken negatively, for example: *Jasmin and George are playing catch. Jasmin drops the ball, George says: 'You are great at this!'*. Previous research shows that the development of irony understanding takes several years. Only when children are around the age of 9, they have reached a full understanding of irony. However, it is still unknown how the development of irony understanding proceeds. Most previous studies on irony used elaborate stories. A disadvantage of elaborate stories is that the participant has to remember everything to answer questions, which requires a lot of cognitive resources. This makes it difficult to test irony understanding, which also requires resources. In this study, we use short video fragments to minimize the cognitive workload. Video fragments including a statement expressing ironic criticism were shown to participants, followed by five questions. Each question addressed a different aspect of irony understanding. In this study 6- to 8-year old children ($N = 70$) and adults ($N = 16$) participated. We found that children are not adult-like in certain aspects of irony understanding. In particular children do not understand the second-order intention of an ironic speaker. The second-order intention of the speaker in the example above was: *'George intends Jasmin to believe that she is bad at catching.'* From our results it is clear that even 8-year old children can not yet fully understand irony. Furthermore we found that children gave different justifications for speaker's non-literal statements than being ironic. Adults say that the speaker is being ironic, while children say mostly that the speaker means the opposite without mentioning the reason why. Moreover, we did not get better results than previous work that used more cognitively demanding materials. Hence, this study has provided a deeper insight into irony understanding and its development of irony understanding.

Keywords: Language acquisition; Theory of Mind; Verbal irony; Ironic criticism; Children, Second-order intention.

Acknowledgements

I would like to thank Professor Petra Hendriks and Iris Scholten, my research supervisors, for their guidance, useful critiques and enthusiastic encouragements on this research work. I would also like to thank my supervisor Dr. Jennifer Spenader, for her advice and feedback during this project, also I want to thank her for her support and assistance in keeping my progress on schedule. My grateful thanks are also extended to Callista Jippes for her valuable technical support on this project and Dr. Jacolien van Rij-Tange for her help on the statistical tests and procedures.

I would also like to thank O.B.S Brinkschool, who provided a place for me to test the children and helped me contact the parents. I am thankful for the participation of the children and the parents consent. The experiments went smoothly, because of all the help I got from the school. Moreover, I want to thank the actors and speakers for their help in this project.

Furthermore, I am thankful for the years of valuable friendship with Kim Veltman, and the encouragement I got from Everhard Tilstra last year. I am particularly grateful for the personal assistance given by Klaske Willy and Savannah, and their wonderful guidance throughout the last couple of years. Finally, I wish to thank my parents and family for their support and encouragement throughout my study.

Contents

1 Introduction **1**

2 Background **3**
2.1 Irony . . . 3
2.2 Theory of Mind . . . 4
2.3 Development of Irony Understanding . . . 6
2.4 Summary . . . 12

3 Research Questions and Hypotheses **15**

4 Method **19**
4.1 Participants . . . 19
4.2 Materials . . . 20
4.2.1 Scenarios . . . 20
4.2.2 Literal and Ironic Statement . . . 20
4.2.3 Questions . . . 21
4.2.4 Coding and Scoring . . . 24
4.3 Design . . . 25
4.4 Video Fragments . . . 25
4.4.1 Actors . . . 25
4.4.2 Green Screen . . . 26
4.4.3 Recording Procedure . . . 27
4.4.4 Technical Details . . . 27
4.5 Audio Fragments . . . 27
4.6 Procedure . . . 28

5 Results **29**
5.1 Meaning Question . . . 29
5.1.1 Children versus Adults . . . 30
5.1.2 6-y.o. versus 7-y.o. versus 8-y.o. . . . 31
5.2 Belief Question . . . 32
5.2.1 Children versus Adults . . . 32
5.2.2 6-y.o. versus 7-y.o. versus 8-y.o. . . . 33

5.3 Emotion Question . . . 35
5.3.1 Children versus Adults . . . 35
5.3.2 6-y.o. versus 7-y.o. versus 8-y.o. . . . 36
5.4 Intention Question . . . 37
5.4.1 Children versus Adults . . . 38
5.4.2 6-y.o. versus 7-y.o. versus 8-y.o. . . . 41
5.5 Justification Question . . . 42
5.5.1 Children versus Adults . . . 43
5.5.2 7-y.o. versus 8-y.o. . . . 45
5.6 Relation between Intention and Justification . . . 47

6 Discussion **49**
6.1 Findings . . . 49
6.2 Limitations and Future Work . . . 51
6.3 Larger implications . . . 53

7 Conclusion **55**

References **59**

Appendices **63**
A Practice and Test Items . . . 63
B Video Fragments . . . 72
C Experiment Lists . . . 73
D Categories of Justifications . . . 74

1 Introduction

Communication is not limited to understanding the literal meaning of the words used, but also includes understanding a lot of other aspects. These aspects encompass the meaning of the words in context, their metaphorical meaning, body language, intonation and prosodic features in speech.

In some utterances the message has a deeper meaning, for example in sentence (1.1), the message is not that the main color of the neighborhood is green. In this statement, the word 'green' represents nature and therefore the intented message of this sentence is that there is a lot of trees in this neighborhood. In sentence(1.1), the statement has a metaphorical meaning.

(1.1) This neighborhood is very green.

In some statements the message has an additional or different meaning than the literal statement. In irony, the intended message often is the opposite of what is literally said. An example of irony is shown in situation (1.2).

(1.2) It is a grey and rainy day.

a. *Ironic statement:* 'What fantastic weather!'
b. *Intended message:* 'What horrible weather!'

We know from previous studies that even at 8-years old, children do not fully understand verbal irony. Children develop irony understanding between 6 and 8 years old (Ackerman, 1981; Pexman & Glenwright, 2007; Filippova & Astington, 2008). However, it is still unclear how children develop irony understanding. The aim of this study is to get more information about the irony understanding of typically developing children. Previous studies suffer from some limitations, for example elaborate stories, which may have made the task too demanding cognitively. To address these limitations, we developed and carried out an experiment that should minimizes these limitations. We investigate whether 6- to 8-year old children will understand irony if the task is less cognitive demanding.

2 Background

2.1 Irony

Verbal irony is a form of non-literal language (Winner, 1997). In irony, the literal statement is not congruent with the meaning of a message. Gibbs (2000) inspected irony usage and identified that verbal irony includes multiple types, such as: jocularity, sarcasm, hyperbole, rhetorical questions, and understatements. According to Kumon-Nakamura et al. (1995), the most common form of verbal irony is when the statement is the opposite of the intended meaning. This counterfactual verbal irony includes two types: ironic compliments and ironic criticism. In ironic compliments, example (2.1) the intended meaning is positive even though the literal statement negative. In ironic criticism, example (2.2), the intended meaning of the message is negative even though the literal statement is positive. Ironic criticism is also known as sarcasm and according to Capelli et al. (1990) it is frequently used.

(2.1) Jane throws the basketball, scoring the hoop.
George says: 'You did great!' *Literal compliment*
George says: 'You did terrible!' *Ironic compliment*

(2.2) Jane throws the basketball, missing the hoop.
George says: 'You did great!' *Ironic criticism*
George says: 'You did terrible!' *Literal criticism*

An interesting part of irony is that it has almost always a victim (Sperber & Wilson, 1981); a target to which the statement is aimed for. Gibbs (2000) found that the victim of the irony could be different things, for example: a situation, a past event, the addressee, or the speaker. The victim in examples (2.2) and (2.1) is a person, namely Jane.

Understanding irony is complex, because multiple components need to be understood (Capelli

et al., 1990). These components are illustrated in (2.3). First of all, the listener has to understand that the literal statement (2.3-a) is not the same as the belief of the speaker (2.3-b). Furthermore, the listener would need to understand the intention of the speaker (2.3-c), e.g. what does the speaker want the listener to belief?

(2.3) Helen (20-y.o.) and Ruben are playing catch. Helen drops the ball often. After the game, Ruben makes an ironic statement.

a. *Statement* Ruben says: 'Wow, you are good at catching!'
b. *Belief* Ruben thinks Helen is bad at playing catch.
c. *Intention* Ruben intends Helen to know that he thinks that Helen is bad at playing catch.

In conclusion, irony gives a statement an additional meaning; the statement does not only contain the literal meaning anymore. Given the complexity of irony, to understand it requires a lot of cognitive resources. It is not surprising that children have difficulty understanding it. Even adults are not always able to recognize irony and sometimes need to verify with: 'Are you being ironic?' or 'Do you really mean this?'

Not only cognitive resources are required to understand irony. It is also important to be able to reason about someone else's mental state, such as understanding the speaker's intentions (that they do not mean what they say). This requires second-order Theory of Mind, which is reasoning about one's mental state is called. Before continuing with the development of irony understanding ToM will be discussed.

2.2 Theory of Mind

As mentioned above, to understand irony the listener needs to understand the speaker's intention, which requires Theory of Mind. Theory of Mind (ToM; Woodruff et al. (1978); Wimmer & Perner (1983); Perner & Wimmer (1985)) is the ability to reason about the mental states of others to communicate and interact in society. These mental states could be beliefs, knowledge, intentions, desires, and opinions of yourself or someone else. According to Shamay-Tsoory & Aharon-Peretz (2007) and Poletti et al. (2011) ToM encompass two kinds, namely; *cognitive ToM* (knowledge and beliefs) and *affective ToM* (feelings and emotions). As will be discussed in Chapter 4, we formulated questions for the participance based on both cognitive and affective ToM.

The story in example (2.4) illustrates a situation in which Leo wants to buy a gift for Kirsten. In the story there are three people present and they all have different information. This results in reasoning to different degrees about what knowledge people have. The zeroth-order ToM is non-social reasoning, it is about what an individual knows or believes themselves. Zeroth-order ToM has nothing to do with the beliefs or knowledge of someone else. An example of zeroth-order ToM is that Leo knows when it is Kirsten's birthday or 'I believe x'. The first-order ToM is the capacity of reasoning about the mental state (feelings/knowledge/beliefs) of someone else's: 'I know that P believes x'. Second-order ToM is 'I know that P2 knows that P1 believes x'.

(2.4) Leo and Kirsten are best friends. Leo knows when Kirsten's birthday is and he is talking to his mom about the present he will give Kirsten.
0th-order Leo knows: It is Kirsten's birthday tomorrow.
1st-order Kirsten knows: Leo is aware that my birthday is tomorrow and I do not know what gift I will get.
Leo discusses with his mom what to buy Kirsten.
2nd-order Mother knows: Kirsten does not know what Leo will give her.

The development of ToM reasoning proceeds step by step. Before the age of four typically developing (TD) children make interpretations based on their own knowledge and perspective (Wellman et al., 2001). Around the age of four, children begin to understand the first-order ToM (Wimmer & Perner, 1983; Wimmer et al., 1984). Second-order ToM reasoning is a complex skill and it will take some more years to develop it. Children start to develop second-order reasoning between 6- to 8-years old (Perner & Wimmer, 1985).

Previous studies have explored the relationships between ToM development and irony understanding. Second-order intention is about what the speaker (1st-order) wants the listener to believe (2nd-order), and this is important to understand irony (Winner & Leekam, 1991; Pexman & Glenwright, 2007). Multiple studies show that children who fail on a second-order ToM task will be more likely to fail to understand irony (Winner & Leekam, 1991; Happé, 1995; Filippova & Astington, 2008).

Understanding the second-order intention is crucial to understand irony, however it is also crucial to understand white lies. Winner & Leekam (1991) investigated the difference between irony and a (white) lie. A difference between a white lie and ironic statement is the second-order intention of the speaker. A liar does not want his true belief to be noticed. The liar wants to cover their, often negative, opinion and (s)he does not want that the listener finds out that (s)he does not say what (s)he really thinks. An ironic speaker wants his true belief to be noticed. The difference between a white lie and an ironic statement is pointed out in the example in Table 2.1.

Table 2.1: Difference in intention between a white lie and an ironic statement.

	White lie	**Ironic statement**
Situation	Anna (5-y.o.) and her brother are playing catch. Anna drops the ball often. After the game, her brother says:	Rosa (20-y.o.) and her brother are playing catch. Rosa drops the ball often. After the game, her brother makes an ironic statement.
Statement	'Wow, you are good at catching!'	
Belief *1st-order ToM*	The brother thinks Anna is bad at playing catch.	The brother thinks Rosa is bad at playing catch.
Intention *2nd-order ToM*	The brother does not intend Anna to know that he thinks that Anna is bad at playing catch.	The brother intends Rosa to know that he thinks that Rosa is bad at playing catch.

As seen in Table 2.1, the intention of the speaker is important to distinguish an ironic statement from a white lie. Furthermore, in irony the listener has to be aware of the non-literal intention of the speaker otherwise the listener could accuse the speaker of saying a white lie. To clarify, imagine that Rosa from Table 2.1 does not realize that her brother intends Rosa to know that he thinks Rosa is bad at playing catch. In that case, Rosa is not getting that her brother is being ironic, even if she realizes that the brother is not saying what he thinks. Therefore, she could accuse him from lying.

Besides this intentional difference, some researchers also argue that the attitude of a liar and an ironic speaker differ (e.g. Winner & Leekam (1991). With an ironic statement, the speaker intends to present a negative attitude, while a liar intends to present a positive attitude. In the study of Winner & Leekam, the participants were presented with two stories. In one story the final utterance was intended to be ironic and in the other the final utterance was intended to be a deception. According to Winner & Leekam the ability to distinguish the intentions of an ironist versus a liar is needed to be able to distinguish the attitude of an ironist versus a liar.

2.3 Development of Irony Understanding

The development of irony understanding takes time. Ackerman (1981) found that young children (first graders') have been shown to interpret utterances as sincere, even when the statement is not congruent with the situation. Children between 6- to 9-years old begin to understand aspects of ironic intent (Pexman & Glenwright, 2007; Filippova & Astington, 2008).

Filippova & Astington (2008) compared advanced social reasoning, which is revealed in irony comprehension, from 5-, 7- and 9-year old children ($N = 72$) and adults ($N = 24$) in a series of experiments. In one of the tasks of this study, different stories were read to the participants accompanied by simple illustrations. After each story the participants had to answer a set of questions. These questions measure the participant's ability to make judgments about someone else's mental state. Figure 2.1 is an item from this task. The *comprehension questions* are zeroth-order ToM questions. The *meaning* and *belief* questions are first-order ToM questions. The *communicative intention* question is a third-order ToM question. Lastly, the *motivation/attitude* question is a second-order ToM question.

Filippova & Astington found that although 7- and 9-year old children are not adult-like in irony comprehension, they are better than 5-year old children. This study indicates a developmental pattern in the ability to interpret ironic discourse. The listener needs to understand the speaker's belief, the difference between the intended and the expressed meaning of the statement, in order to understand the intention of a speaker. Furthermore, these three aspects are needed for someone to understand the speaker's attitude. Additionally, Filippova & Astington found that not only age and memory contributes to children's understanding of irony but also ToM and language ability.

Filippova & Astington used elaborate stories in their experiment. A disadvantage is that the participant has to remember a lot to answer questions; for instance names, situations, and actions. Take for example the story in Figure 2.1, the participant has to remember for example the boys names, that they are best friends, that Robert is new on the team, that Robert is excited about being on the team, that Robert misses several goals, the utterance. To remember all this requires a lot of cognitive resources. It is not clear if the demands of the task might have made it harder for children

Robert is the new player on his school's soccer team. He is really excited about being on the team. Robert's best friend Oliver also plays on the team. During his first game, Robert misses the chance to score several easy goals. His team loses the game. After the game, Oliver says to Robert: 'You sure ARE a GREAT scorer!'

Comprehension Questions	1)	Did Robert help his team to win the game?
	2)	What did Oliver say to Robert?
Meaning	3.1)	Does Oliver mean that?
	3.2)	What does he mean?
Belief	4)	Does Oliver think Robert is a great scorer?
		Oliver said to Robert: 'You sure ARE a GREAT scorer!
Communicative intention	5)	Does Oliver want Robert to believe that he thinks that?
		Oliver said to Robert: 'You sure ARE a GREAT scorer!'
Motivation/Attitude	6)	Why did he say that?

Figure 2.1: Item from Filippova & Astington (2008) page 137.

to be successful in recognizing others' mental state. Two ways to lower the cognitive demands for the task could be to use short stories or video fragments. In a shorter story the participant does not have to remember unnecessary details, for example the names of the characters. In a video fragment the participant will see the actions and therefore the participant does not have to make an internal visualization of the story. Both solutions will lower the cognitive resources needed to remember the story. Therefore, the participants are able to use their cognitive resources to focus on the task.

Even though most experiments investigating irony consist only of close-ended questions, Filippova & Astington's experiment includes an open-ended question (question 6 in Figure 2.1). This open-ended question about the motivation of the speaker is very informative, because it gives insight into what the participant thinks the motivation of an ironic speaker is. Thereby giving more information about the development in reasoning than a simple 'YES'-'NO' question. Filippova & Astington used a scoring system to assigns a score depending on the ToM level of the answer. The responses were coded on a 4-point scale, with a higher score for a higher cognitive complexity of the responses. A score of 0 was given for responses which were irrelevant; a score of 1 was given for responses which reflect a simple surface-level justification; a score of 2 was given for responses which reflect identification of the speaker's first-order intention; a score of 3 was given for responses which reflect identification of the speaker's second-order intention; and a score of 4 was given to responses which reflect identification of an attitude. A limitation with this scoring system is that a response was given a full 4-point score when it mentioned the speaker's attitude, regardless the absent of the indication that the speaker was being ironic.

Hancock et al. (2000) obtained similar results to Filippova & Astington. They researched with two experiments 5- and 6-year old children's development of understanding non-literal speech (Experiment 1: $N = 24$, Experiment 2: $N = 24$). Participants watched short videotaped stories with an ironic compliment, ironic criticism, literal compliment, or literal criticism. After each story, three questions were posed. The first question was a first-order belief question, similar to the *belief* question from Filippova & Astington. The second question asked whether the speaker was being nice or mean. Hancock et al. calls this *speaker's intent* and it requires second-order ToM reasoning. However to prevent confusion we will call it *speaker's attitude*, because other researchers also call this attitude (e.g. Pexman & Glenwright, 2007). The last question was about the beliefs of the participant, which is a zeroth-order ToM question. Figure 2.2 shows an item from experiment 1 of Hancock et al..

The Weight Lifter Story

Critical version

A: I'm good at lifting weights.
[A fails to lift the weights.]
B: You are bad at lifting weights. *Literal condition*
or
You really are good at lifting weights. *Ironic condition*

Complimentary version

A: I'm good at lifting weights.
[A lifts the weights.]
B: You are good at lifting weights. *Literal condition*
or
You really are bad at lifting weights. *Ironic condition*

Questions:

1. *Speaker's belief* Did B really think that A was a good weight lifter?
2. *Speaker's attitude* Was B being mean or nice?
3. *Participant's belief* Do you think A was a good weight lifter?

Figure 2.2: Item from Experiment 1 from Hancock et al. (2000) page 231. This item is adapted for readability and questions are include.

Hancock et al.'s results indicate that the participants found it easier to recognize that the speaker was being non-literal (38%) than to interpret speaker's attitude (28%). This is in line with the findings from Filippova & Astington (2008); although children detect the non-literal meaning of an ironic statement, they are not yet able to recognize irony. Another finding from Hancock et al. is that children are asymmetrical across detecting ironic criticisms and ironic compliments.

Unlike Filippova & Astington (2008), Hancock et al. used short videotaped stories. Short video fragments are a very reliable method because the participant has to remember fewer facts. In the story in Figure 2.2 the participant only has to remember that someone fails to lift the weights and

the utterance. Therefore the cognitive workload is lower than with an experiment with a long story, which makes that the participant can use much more cognitive resources to perform the task. However, the statement character A makes in Figure 2.2 is redundant and could be left out.

Nevertheless, a downside of Hancock et al.'s experiment is that the ironic sentences contain the word 'really', while the literal sentences do not include this. This could be a confound in their experiment. Furthermore, Hancock et al. do not have a question about the speaker's intention for example: 'Did B want A to believe him?'.

Another study about ironic compliments and ironic criticism was done by Pexman & Glenwright (2007). They did research on how children develop the ability to understand verbal irony. Typically developing children (6- to 10-year olds, $N = 70$) were presented short prerecorded narratives puppet shows with four different endings: literal compliment, ironic compliment, ironic criticism, or literal criticism. After each show, the participants had to answer three questions. The first question is a first-order belief question (similar to *belief* from Filippova & Astington (2008), and the first question from Hancock et al. (2000)), with two possible answers. The second question is an second-order attitude question (similar to the second question of Hancock et al., 2000) and the participant had to say on a 5-point scale how nice or mean the speaker was. The third question is a second-order teasing question, and the participant had to answer on 3-point scale how teasing the speaker was. Figure 2.3 is an item of this study.

Sample puppet show

This is Sam, and this is John. They play on the same soccer team. It is the last few minutes of a game.

Outcome 1:

John kicks the ball, scoring a goal.

1.1	Sam says: 'That was a great play!'	*Literal compliment*
1.2	Sam says: 'That was a terrible play!'	*Ironic compliment*

Outcome 2:

John kicks the ball, missing a goal.

2.1	Sam says: 'That was a great play!'	*Ironic criticism*
2.2	Sam says: 'That was a terrible play!'	*Literal criticism*

Questions:

1. *Speaker Belief Question:* When Sam said 'That was a great play!', did he think John made a good or a bad play?
2. *Speaker Attitude Question:* Point to one of the faces to show me how nice or mean Sam was trying to be when he said, 'That was a great play!'.
3. *Speaker Teasing Question:* Point to one of the faces to show me how teasing or real Sam was trying to be when he said, 'That was a great play!'.

Figure 2.3: Item from Experiment 1 from Pexman & Glenwright (2007) page 231. Adapted for readability and included questions.

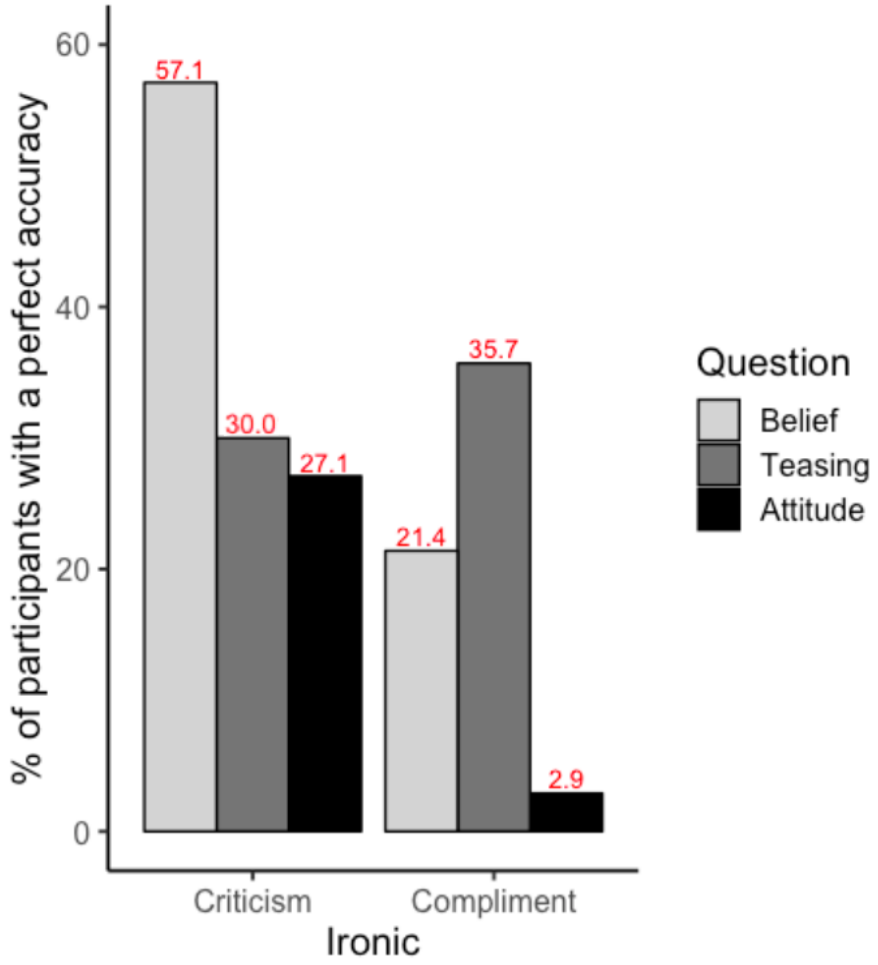

Figure 2.4: Visualization of the results from Pexman & Glenwright (2007).

The results of Pexman & Glenwright are visualized in Figure 2.4. They found that for the ironic criticism items, significantly more participants had perfect accuracy for the *speakers belief* question than for the *speakers teasing* question. For the ironic compliment items, significantly more participants had perfect accuracy for the *speakers teasing* question than for the *speakers attitude* question. This leads to their conclusion that the developmental progression of ironic criticisms and ironic compliments are different. For ironic criticism, first the understanding of the speaker's belief will emerge. Followed by together the understanding of the speaker's intent (*teasing question*) and speaker's attitude. Concerning ironic compliments, the understanding of the speaker's belief and the understanding of the speaker's intent (*teasing question*) emerge together, followed by the understanding of the speakers attitude.

It can also be seen from Figure 2.4 that for the speaker's belief and speaker's attitude question, more children had a perfect accuracy for ironic criticism than for ironic compliment for speaker belief judgment. This shows that children have more difficulty understanding ironic compliments than ironic criticisms. Pexman & Glenwright mention that a possible explanation could be that

ironic criticisms occur more often than ironic compliments. The finding that children understand ironic criticism and ironic compliments differently is in line with the findings of Hancock et al. (2000).

An advantage of this puppet show is that it is visual and short. Still, there is room for improvement, namely actors could act out the situation instead of a little story to decrease the amount of sentences. Furthermore, this experiment does not contain a question about the intention of the speaker. Hence, a more full experiment about irony understanding should include such a question.

Not only children have difficulties with understanding irony. One example of this is shown by Pell et al. (2014). They did research on the ability of adopting theory of mind with nondemented patients with Parkinson's disease (PD). A sub-part of the experiment investigated how well participants could interpret enriched social intentions such as sarcasm or lies. Adults with PD (N=15) and healthy adults (N=16) participated in this experiment about interpretation of emotions and social intentions in communication. Pell et al. made use of *The Awareness of Social Inference Test* (TASIT; McDonald et al. (2003). This test consists of video vignettes of everyday social interactions performed by professional actors in English. One part (Part 3 of TASIT) of the experiment focuses on enriched social intentions. Figure 2.5 is an example of an item of Part 3 of TASIT. The *do* question requires second-order ToM reasoning and is similar to *Communicative intention* question. The *say* question is a first-order ToM question and is similar to the second *meaning* question from Filippova & Astington (2008). The first-order *think* question is similar to *belief* from Filippova & Astington, the first question from Hancock et al. (2000) and *speaker belief* from Pexman & Glenwright (2007). The *feel* question is in this item requires second-order ToM reasoning. However, there are also items in TASIT in which the *feel* question is a first-order ToM question.

Tanya (carrying a plate):	Cal would like some ice cream
Mick (busy, not looking at plate):	As long as he has eaten all his dinner.
Tanya:	Oh, yeah, he's eaten all his dinner, all right.
(Camera shows close-up of nearly full plate - hidden from Mick in 'lie' mode but visible to him in 'sarcasm' mode.)	
Mick:	I told him to finish everything on his plate.
Tanya:	Well, he certainly listened to you.
Probe questions	
Do	Is Tanya trying to make Mick believe Cal ate all his dinner?
Say	Does she mean that Cal obeyed Mick?
Think	Does Mick think Cal ate all his dinner?
Feel	Is Tanya hoping Mick will think Cal at his dinner?

Figure 2.5: Item Part 3 TASIT (Pell et al., 2014) page 5.

Pell et al. (2014) found that mild-moderate PD patients showed deficiencies in their ability to interpret enriched social intentions such as sarcasm or lies from non-literal remarks.

On the one hand, the *feel* question in Pell et al.'s experiment is not seen in the other experiments. This is a question which is interesting because it shows whether participants are able to reason at affective ToM. On the other hand, a downside of this experiment is that the video fragment is quite

long. It takes multiple sentences and the participant has to take in all information to remember it. As mentioned earlier, this will take a lot of cognitive resources and therefore less resources are available to focus on the task.

2.4 Summary

To understand irony it is important to be able to reason about someone else's mental state, which is called Theory of Mind (ToM). In particular understanding speaker's intentions is important to understand irony. This requires second-order Theory of Mind. According to Perner & Wimmer (1985), children start to develop second-order ToM reasoning between 6- to 8-years old. Furthermore, given the complexity of irony, understanding irony also requires a lot of cognitive resources. It is widely known that children do not have fully developed cognitive resources.

Filippova & Astington (2008) show that 5-, 7-, and 9-year old children are not adult like in irony understanding and Pexman & Glenwright (2007) show that only around half of the 6- to 10-year old children have a perfect accuracy for questions related to irony understanding. This shows that although children of 9- and 10-years old are able to reason in second-order ToM, they are not adult-like in understanding irony. One reason for these results could be that these previous experiments demanded a lot of cognitive resources, which made it harder for children to be successful in recognizing other mental states.

In some experiments, the participant had to listen to an elaborate story (Filippova & Astington, 2008; Winner & Leekam, 1991). A disadvantage of using elaborate stories is that the participant has to remember the whole story to answer the questions. This requires a lot of cognitive resources, and therefore there are less cognitive resources available to process an ironic statement. A more comprehensive study would use stories with as little information as possible.

In an irony study conducted by Pexman & Glenwright (2007), participant watched a puppet show. The story in this experiment is substantially shorter than in Filippova & Astington and Winner & Leekam. However, still it contains names and other details the participant has to remember requiring more cognitive resources

Pell et al. (2014) used The Awareness of Social Inference Test (TASIT; designed by McDonald et al. (2003)) to find that mild-moderate Parkinson's disease patients showed deficiencies in their ability to interpret enriched social intentions such as sarcasm or lies from non-literal remarks. This test contains videotaped vignettes. Nevertheless, the vignettes were a conversation with multiple sentences. Therefore, it could be even more minimized.

Hancock et al. (2000) also used video taped stories. In these stories a short situation is acted out, followed by two sentences. The reduced conversation in this experiment is very useful, although it could be even more reduced.

Overall, some of the questions in the mentioned studies are similar to each other. An overview of the related questions is shown in Table 2.2. In these examples, only Filippova & Astington (2008) used a open-ended question in their method. In contrast to a 'YES'-'NO' question, an open question gives a lot of insight in what a participant thinks.

Table 2.2: Questions from previous studies.

Filippova & Astington (2008)	Hancock et al. (2000)	Pexman & Glenwright (2007)	Pell et al. (2014)
Meaning			Say
Belief	Speaker's belief	Speaker belief	Think
			Feel
Communicative intention			Do
Motivation / Attitude			
Comprehension			
	Speaker's attitude	Speaker attitude	
	Participant's belief		
		Speaker teasing	

In light of what has been discussed in this chapter, we are interested how children will perform in an experiment which requires less cognitive resources than previous studies. In order to decrease the complexity of the task, participants have to remember less information and there will be more resources available to process an ironic statement. Furthermore we are interested in the responses of children on an open-ended justification for the speaker's non-literal statement.

3

Research Questions and Hypotheses

It has previously been observed that children start to develop second-order Theory of Mind (ToM) reasoning between 6- to 8-years old (Perner & Wimmer, 1985). Being able to reason at second-order ToM is important to understand irony because one has to understand the second-order intention of the ironic speaker. Furthermore, understanding irony also requires a lot of cognitive resources. Previous research has established that children of 9- to 10-years old are not adult-like in understanding irony (e.g. Filippova & Astington (2008); Pexman & Glenwright (2007)). These results combined show that even though children of 9- and 10-years old are able to reason in second-order ToM, they do not understand irony. A possible explanation for these contradictory results may be that these previous studies on irony used elaborate stories. A disadvantage of elaborate stories is that the participant has to remember the whole story in order to answer the questions, which requires a lot of cognitive resources. This makes that the participant does not have all resources available to understand the ironic statement. In view of all that has been discussed so far, one may suppose that if the complexity of the task is reduced, children will show irony understanding at an earlier age. In this study, we investigate the irony understanding of children between 6- to 8-year old children when the task is less complex. To reduce the cognitive workload for the task, we use short video fragments.

The central interest in this study is to investigate the development of irony understanding in typically developing children. In order to assess this, 6-, 7-, and 8-year old children are compared with adults (Research question 1), and with each other (Research question 2). To fully understand irony, someone has to understand multiple components related to different orders ToM. We formulated two research questions with five sub-questions. Each sub-question focuses on a different component which is needed to understand irony. For the first research question we will compare children with adults. The first research question, sub-research questions and the hypothesis are stated below.

Research Question 1: Can 6- to 8-year old children show adult-like irony understanding ?

1a. Can 6- to 8-year old children show adult-like understanding that the meaning of an ironic statement has to be taken non-literal?

1b. Can 6- to 8-year old children show adult-like judgment of the first-order belief of an ironic speaker?

1c. Can 6- to 8-year old children show adult-like identification of emotions expressed by an ironic speaker?

1d. Can 6- to 8-year old children show adult-like understanding of the second-order intention of an ironic speaker?

1e. Can 6- to 8-year old children show adult-like justifications for the speaker's non-literal statement?

Young children are able to detect that an ironic speaker is making a non-literal statement (Ackerman, 1983; Hancock et al., 2000; Filippova & Astington, 2008). Based on this information our expectation for sub-research question 1a is that 6- to 8-year old children are adult-like in understanding that an ironic statement has to be taken non-literally.

Children around the age of 4 are able to reason at a first-order ToM level (Wimmer & Perner, 1983; Wimmer et al., 1984). This leads to our expectation for sub-research question 1b that 6- to 8-year old children are adult-like in judging the first-order belief of an ironic speaker.

Sub-question 1c is about the emotion expressed by an ironic speaker. Emotion recognition starts to develop around 3- or 4-year old Pons et al. (2004). Therefore we expect that the children will be adult-like in identifying emotions.

Previous research has shown that children are actually not adult-like in understanding the second-order intention of an ironic speaker (Ackerman, 1983; Hancock et al., 2000; Filippova & Astington, 2008). Even though with a reduced complexity of the task, our expectation for sub-question 1d is that children between 6- to 8-years old are not completely adult-like in understanding the second-order intention of an ironic speaker. Furthermore, our expectation for sub-question 1e is that children between 6- and 8-years old are not adult-like in their justification for the speaker's non-literal statement.

Hypothesis 1: Children from 6- to 8-year old are not fully adult-like at irony understanding. They are adult-like in understanding that the meaning of an ironic statement has to be taken non-literal, judging the first-order belief of an ironic speaker, and identifying the emotion expressed by an ironic speaker. However, they have difficulty with understanding the second-order intention of an ironic speaker and their justifications for this are not adult-like.

We formulated the second research question to investigate the development of irony understanding in typically developing children between the ages of 6-8. In order to assess this, 6-, 7-, and 8-year old children are compared with each other. Again, the research question contains five sub-questions.

Each one focuses on a different component which is needed to understand irony. The second research question, sub-research questions and the hypothesis are stated below.

Research Question 2: Do 6- to 8-year old children develop with age at irony understanding?

2a. Do 6- to 8-year old children develop with age in understanding that the meaning of an ironic statement has to be taken non-literal?

2b. Do 6- to 8-year old children develop with age in judging the first-order belief of an ironic speaker?

2c. Do 6- to 8-year old children develop with age in identifying emotions expressed by an ironic speaker?

2d. Do 6- to 8-year old children develop with age in understanding the second-order intention of an ironic speaker?

2e. Do 6- to 8-year old children develop with age in their justification for the speaker's non-literal statement?

As mentioned earlier, we think children at age 6- to 8-years will be able to understand that the meaning of an ironic statement has to be taken non-literal, judge speaker's first-order belief, and identify emotions of an ironic speaker. Therefore, we expect that younger children perform similar on these aspects compered to older children. However due to decreased complexity of the task, we expect that older children perform more adult-like in understanding the second-order intention of an ironic speaker and the justification of using non-literal statements.

Hypothesis 2: There will be a development of irony understanding present among children from 6- to 8-years old. The children will be similar in understanding that the meaning of an ironic statement has to be taken non-literal, judging the first-order belief of an ironic speaker, and identifying the emotion expressed by an ironic speaker. However, older children are more adult-like than younger children in understanding the second-order intention of an ironic speaker and their justifications for this are more adult-like.

In the experiment participants will be shown short video fragments including an ironic statement. The video's were followed by five questions. The five questions in the experiment are each focused on one sub-research question. In our experiment we focus on ironic criticism because that is to be believed the easiest form of ironic (Demorest et al., 1983; Kumon-Nakamura et al., 1995; Hancock et al., 2000; Pexman & Glenwright, 2007).

4
Method

The present study was designed to determine whether children can be adult-like in irony understanding and if children develop with age at irony understanding. The experiment consists of video fragments containing a short scenario acted out by two actors. One of the actors makes a statement about the situation. Each video was followed by four or five questions. Children between 6- and 8-year old (N=70) and adults (N=16) participated in this experiment.

4.1 Participants

Seventy typically developing Dutch children (mean age = 7;9, age range = 6;1 - 8;11, 38 girls) participated in the experiment. They were from the same primary school in the north of the Netherlands. The children's caregivers gave passive consent and the children were told that they could quit participating at anytime. After the experiment the children received a sticker for participating. The children were divided into three different age groups; 6-, 7-, and 8-year old's. We excluded one child from the results because this child was the only participant who was 9 years old (9;3 years). For more information about the participants see Table 4.1.

Sixteen native Dutch speaking adults (mean age = 23;10, range age = 20;6 - 27;5, 9 woman) also participated. Adults gave written informed consent prior to the experiment.

The participants did not have any background information about the purpose of the experiment. The study was approved by the Research Ethics Review Committee (CETO) of the University of Groningen

Table 4.1: Information about the age groups.

Age group	n total	n Female	Range age	Mean age
6-years old (Child)	11	7	6;1 - 6;9	6;4
7-years old (Child)	24	14	7;0 - 7;11	7;5
8-years old (Child)	35	17	8;0 - 8;11	8;5
Adults	16	9	20;6 - 27;5	23;10

4.2 Materials

4.2.1 Scenarios

The scenarios in the video fragments are every day life scenarios which children would recognize, similar to the experiments conducted by Pexman & Glenwright (2007) and Hancock et al. (2000). In the situations something is broken, missing or something goes slightly wrong; for example a broken toy or someone spilling coffee. The scenarios were short and clear, without a long story line. The scenarios had to be clearly visual and therefore could not be about a smell or taste.

In the fragments, one of the two actors makes a statement. The statement is about the situation (4.1-a) or about something else, like an object (4.1-b). In 8 of the 16 test items, the victim was an object and in the other 8 test items the victim was a person. Appendix B shows the victim in each scenario.

(4.1) a. *Person is victim of irony (Man)*
The man and woman are sitting at a table. The man is cutting a pie. He gives the woman a small piece and himself a really big one. The woman says: 'Gee, you divided the pie equally!'

b. *Situation is victim of irony (Table)*
The man and woman are sitting at a very messy table. The man says: 'Wow, this is tidy.'

In some of the scenarios an object was needed, such as a pie or a sandwich. It was clearly visible for the participant. Likewise, in some of the scenarios a special background was needed; a lot of snow or a broken swing.

In the case of the two practice scenarios, there was a neutral scenario with a neutral utterance.

4.2.2 Literal and Ironic Statement

Every video fragment contains a statement in Dutch, which was a short sentence with an opinion about the scenario. For example; 'That was a horrible handstand'.

Because ironic criticism is the most common and easy type of irony (Demorest et al., 1983; Kumon-Nakamura et al., 1995; Hancock et al., 2000; Pexman & Glenwright, 2007), we used ironic

criticism and literal criticism in our experiment.

For every scenario two versions were recorded. Both recordings contained the same situation, but the utterance was different. The first was with a literal utterance and the second was with a ironic utterance. In the literal criticism, the sentence contains negative-evaluation-word ('bad', 'horrible') about the scenario, in which the sentence is congruent with the situation. Contrary, the ironic criticism contained a sentence with a positive-evaluation-word ('good', 'fantastic') about the scenario, where the sentence is incongruent with the situation. Example (4.2) is an item from the experiment, in the literal condition the woman utters sentence (4.2-a) while in the ironic condition the woman utters sentence (4.2-b).

(4.2) A man and woman are sitting at a table. The man is cutting a pie. He gives the woman a small piece and himself a really big one. The woman says:

a. Literal: 'Jeetje, wat heb jij de taart oneerlijk verdeeld!'
'Gee, you divided the pie unequally!'

b. Ironic: 'Jeetje, wat heb jij de taart eerlijk verdeeld!'
'Gee, you divided the pie equally!'

The literal and ironic utterances differ minimally; only the evaluation-word differs (respectively negative and positive). Looking at the pie example, in the literal condition the woman says 'unequally' (sentence (4.2-a)) and this is replaced with 'equally' for the ironic condition (sentence (4.2-b)).

The utterances were short sentences (words per sentence: range=4-8, mean=6.2 (Appendix B). The utterances contained a evaluation about the situation, without a reason or result. Furthermore the statements did not contain a negation, because previous studies have shown that negation is more difficult to process and understand (Cornish & Wason, 1970; Lüdtke et al., 2008; Margolin & Abrams, 2009).

To make the statements more natural, it contains an interjection (a word or expression that expresses a feeling or reaction like 'gee' and 'oh dear'). Kreuz & Caucci (2007) investigated sarcastic statements and their results show that interjections are common in sarcastic statements. Another study that investigated ironic sentences in online comments is conducted by Carvalho et al. (2009). They showed that ironic sentences could be found by exploring clues in comments such as interjections.

Each video fragment was acted out by a man and woman, one of the two actors was the speaker (see Appendix B). For each scenario the speaker and listener were the same in both conditions. Appendix B provides also the length of the fragments (mean length = 8.3 seconds).

4.2.3 Questions

The video fragments were followed by four 'YES'-'NO' questions (*Q-meaning*, *Q-belief*, *Q-emotion*, and *Q-intention*) and one open-ended question (*Q-justification*). The questions are based on the experiments by Hancock et al. (2000); Pexman & Glenwright (2007); Filippova & Astington (2008); Pell et al. (2014). Table 4.2 shows the relation between the questions from these previous studies.

Table 4.2: Questions from current study related to previous studies.

Current Study	Filippova & Astington (2008)	Hancock et al. (2000)	Pexman & Glenwright (2007)	Pell et al. (2014)
Meaning	Meaning			Say
Belief	Belief	Speaker's belief	Speaker belief	Think
Emotion				Feel
Intention	Communicative intention			Do
Justification	Motivation / Attitude			

The questions evaluate several key aspects of irony. Three questions are about first-order ToM; *Q-meaning*, *Q-belief*, and *Q-emotion*. The other two questions are about the second-order ToM; *Q-intention* and *Q-justification*. We did not introduce character names, the actors were referred to as 'man', 'he', 'woman', or 'she'. An example of a test item followed by the questions is presented in Figure 4.1.

Q-meaning gives information about the participant's ability to understand the meaning of a statement. It is about whether the speaker meant what (s)he said or not. In other words, whether the statement has to be taken literally or not. In irony, the exact statement is not congruent with the intended message. It was phrased as: 'Does the woman mean [...].' For balance, *Q-meaning* has two versions; formulated positively ('divided equally'/'fantastic handstand') and negatively ('divided unequally'/'horrible handstand'). It is a first-order ToM question and it is related to knowledge and beliefs (cognitive ToM; Shamay-Tsoory & Aharon-Peretz (2007); Poletti et al. (2011)). This question is based on the first *meaning* question from Filippova & Astington (2008) and the *say* question from Pell et al. (2014) (see Table 4.2).

Q-belief identifies whether the participant is able to make first-order ToM judgments about the belief of the speaker: what does the speaker think? The belief question gives insight into whether the participant is able to recognize the belief of the speaker. This is crucial to know, because the participant will not accuse the speaker of irony if (s)he beliefs that the opinion of the speaker is positive (in a negative situation). This question is phrased as: 'Does the man think that [...].' This question falls into the category of cognitive ToM which includes knowledge and beliefs, according to Shamay-Tsoory & Aharon-Peretz (2007); Poletti et al. (2011). Additionally, Hancock et al. (2000); Pexman & Glenwright (2007); Filippova & Astington (2008); Pell et al. (2014) all have a belief question (see Table 4.2).

Q-emotion is a first-order ToM question about the feelings and emotions of the speaker. The participants had to answer whether the speaker has a specific emotion. An example is: 'Is the man grumpy/happy/irritated?' This question is related to affective ToM (feelings and emotions) (Shamay-Tsoory & Aharon-Peretz, 2007; Poletti et al., 2011). This question is based on the *feel* question from Pell et al. (2014) (see Table 4.2).

Q-intention is about the second-order intention of the speaker: what does the speaker want the listener to know? Does the speaker want the listener to believe the statement or not? The intention of the speaker makes the difference between a lie and an ironic statement. In the experiment the

question is phrased as: 'Does she want the man to believe this?'. To correctly answer the questions, the participant needs to understand the following: (1) What did she (the speaker) say? (2) What does she think? (3) Is what she says the same as what she thinks? (4) If what she says is not the same as she thinks: Does she want the listener to know that it is not the same? This requires second-order ToM judgment. *Q-intention* always includes the exact statement of the video fragment, and it is phrased: 'The man says to the woman [...] Does he want the woman to believe this? Furthermore it is related to cognitive ToM, which is about knowledge and beliefs (Shamay-Tsoory & Aharon-Peretz, 2007; Poletti et al., 2011). Filippova & Astington (2008) have a similar question which they call *communicative intention*, and Pell et al. (2014) have a similar question, called *do* (see Table 4.2).

Q-justification was asked in the items with a ironic utterance. This question is about justification for speaker's non-literal statement: 'Why does (s)he say it?' It was asked to get more insight of what the participant thought was the intention of the speaker, just as *Q-intention*. *Q-intention* is a closed-ended question which only gives information about whether the participant had answered the question correct or not. Therefore, adding an open-ended justification question gives us more information insight inabout the thoughts of the participant. This question requires a second-order ToM judgment, and it is related to cognitive ToM which is about knowledge and beliefs (Shamay-Tsoory & Aharon-Peretz, 2007; Poletti et al., 2011). Filippova & Astington (2008) have a similar question, namely; *motivation/attitude* (see Table 4.2).

The man and woman are sitting at a table. The man is cutting a pie. He gives the woman a small piece and himself a really big one. The woman says:		
Literal statement		Jeetje, wat heb jij de taart oneerlijk verdeeld! *Gee, you divided the pie unequally!*
Ironic statement		Jeetje, wat heb jij de taart eerlijk verdeeld! *Gee, you divided the pie equally!*
Q1	**Intention**	De vrouw zegt tegen de man: `[STATEMENT]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat de man de taart eerlijk heeft verdeeld? *Does the woman mean that the man divided the pie equally?*
	Negative	Bedoelt de vrouw dat de man de taart oneerlijk heeft verdeeld? *Does the woman mean that the man divided the pie unequally?*
Q3	**Justification**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw dat ze genoeg taart heeft gekregen? *Does the woman think that she got enough pie?*
Q5	**Emotion**	Is de vrouw blij met hoe de taart is verdeeld? *Is the woman happy about how the pie is divided?*

Figure 4.1: Item from the experiment.

To make sure that the participant was not receiving new information through the questions, the questions were asked in a specific order. *Q-intention* involves second-order ToM, therefore it is the most

difficult question from the four 'YES'-'NO' questions to answer correctly. The justification question was asked only with the ironic items. The order of the questions is: *Q-intention*, *Q-meaning*, *Q-justification*, *Q-belief*, *Q-emotion*. One item of the experiment is represented in Figure 4.1. A list of all items can be found in Appendix A.

4.2.4 Coding and Scoring

Questions *Q-meaning*, *Q-belief*, *Q-emotion*, and *Q-intention* had to be answered by 'YES' or 'NO'. Questions *Q-meaning*, *Q-belief*, and *Q-emotion* could be asked with a positive meaning/belief/emotion or with a negative meaning/belief/emotion. The answers on these questions indicate the participants interpretation of the utterance; literal or ironic. Table 4.3 presents the questions and the indicated interpretation of participant's answer of an ironic item.

Table 4.3: Participant's answer and the indicated interpretation of the answer of an ironic item.

Positively or negatively formulated question		**Participant's answer and interpretation**	
		YES	NO
Q-meaning	positive	literal	ironic
Q-meaning	negative	ironic	literal
Q-belief	positive	literal	ironic
Q-belief	negative	ironic	literal
Q-emotion	positive	literal	ironic
Q-emotion	negative	ironic	literal
Q-intention		literal	ironic

The *Q-justification* is an open-ended question, which leads to multiple answers. Therefore we grouped the responses based on their content. The distinct categories which were clearly visible in the responses are: Ironic, White lie, Politeness, Opposite, Proverb, Intonation, Humor, No Answer, and Other. In Appendix D are the participant's responses with their category.

For the qualitative analysis of the *Q-justification* we used the categories, however for the quantitative analysis we gave each category a score. In the experiment of Filippova & Astington (2008) the following question was asked: 'How do you know [child's response]?'. Filippova & Astington used a scoring system to assigns a score depending on the ToM level of the answer. A limitation with this scoring system is that a response was given a full 4-point score when it mentioned the speaker's attitude, regardless the absent of the indication that the speaker was being ironic. Therefore, we adapted the scoring system by assigning scores based on the type of justification for the response on the second-order intention. With lower scores given for more literal responses and higher scores for answers that showed a clearer understanding of irony. The highest score was given to an response which indicated that the participant understood that the speaker was being ironic. Table 4.4 represents the category with score and the type of justification.

Table 4.4: Each category with their score and the type of justification.

Category	Score	Type of justification
1a) No Answer 1b) Other 1c) Literal	0	- The participant does not give an answer to the question. - The participants gives an answer which is not categorizable. - The participant says that the speaker tells the truth.
2a) Intonation 2b) Proverb 2c) Humor	1	The participant knows that the statement should not be taken literally, however, the participant does not say anything about the reason.
3) Opposite	2	The participant knows that the speaker means the opposite but does not say anything about the intention.
4a) Politeness 4b) White Lie	3	The participant knows that the speaker means the opposite, however, the participants does think that the intention is something else than ironic.
5) Ironic	4	The participants knows that the speaker means the opposite and that the intention of the speaker is to let the listener know (s)he means the opposite.

4.3 Design

For every situation there were two video fragments; one with a literal criticism and one with an ironic criticism (Section 4.2.2). This leads to a 2x1 design.

Due to the two statement conditions (Literal criticism and Ironic criticism) and the two possible formulations of *Q-meaning*, we made four balanced lists (Appendix C). Each participant received one list which contained 16 test items in total, 8 items with the literal condition and 8 items with the ironic condition.

4.4 Video Fragments

4.4.1 Actors

Seven men and seven women acted out the scenarios. The actors were students (age: min=20, max=28, mean=24.9) and they were friends of one of the researchers. The actors are students because they are young adults and they are closer to the age from the children. Children may react different to young adults than to older people. The same actors appeared in both condition of each item. All actors were native Dutch speakers with a neutral Dutch accent.

The actors were told to look at each other and not towards the camera. A trade-off was made between showing as much of the face and still standing in a natural position. The speaker was told to articulate clearly and to speak slower than normal for optimal quality of the utterance. The listener was told to not give a reaction on the statement, neither verbal nor non-verbal. The actors were told to employ natural changes in intonation during the ironic video fragments. Intonation in the ironic

criticism video fragments diverge from the intonation in the literal criticism fragments. However, ironic intonation differed per actor.

The actors wore basic clothes, in order not to be distracting for the participants. Furthermore, the actors did not have visible tattoos or piercings. The faces were clearly visible, for example so the men shaved, and no one wore a cap. All in all, the actors did not have a special or unordinary appearance.

The actors signed a consent form in which they gave permission to use the video recordings for academic purposes.

4.4.2 Green Screen

The video fragments were recorded in a studio by using a green screen. The background was edited afterwards. We used a studio to control side effects of the environment. In a studio are less background noises and background distractions, which could have an effect on the quality of the fragments and therefore on the reliability of measurements. Moreover, the practical issue of moving to other locations, which are suited for acting out the scenario, was solved by recording in a studio.

The video fragments were shot in landscape position, therefore we searched for landscape sized background pictures on Flickr[1] and Shutterstock[2]. Each scenario has a unique and appropriate background. For example, the actors played tennis outside (Figure 4.2) and sat in a living room when they drank coffee (Figure 4.3). In some fragments (such as making a handstand), the floor was also a green screen which was edited as a floor in the fragments.

Figure 4.2: Green screen background for the tennis scenario.

Figure 4.3: Green screen background for the coffee scenario.

The background was stationary and serene, without a lot of decoration, to minimize possible distractions. The perspective of the fragments is frontal, hence the perspective of the background had to be frontal as well.

[1] https://www.flickr.com/
[2] https://www.shutterstock.com/nl/

Only visual background was used. While audio background would make the fragment more realistic, it may also add unnecessary distraction. It would also increase the difference between the scenarios, which is not desired.

4.4.3 Recording Procedure

For this experiment we needed 18 different scenarios. We initially shot 26 scenarios and chose the 18 scenarios that were the most clear. We shot every scenario at least 3 times per condition. We selected the best fragments based on the overall quality, such as: visibility of the action, audibility of the utterance, and the intonation of the utterance.

For every scenario, we recorded first the literal condition followed by the ironic condition in order to avoid that the actors were biased to speak ironically. We used separate moments to shoot the video fragments. Firstly, we made pilot fragments to decide what kind of actions were visible and what noticeable should be used as materials for the experiment. Initially a second moment was scheduled to shoot all the fragments. However, after this shooting day we evaluated the recorded fragments and we decided to redo the shooting of some scenarios due to all kinds of mistakes in the fragments. The fragments which we used in the experiment consisted of one scenario from the pilot shooting, thirteen scenarios from the second shooting, and 4 scenarios from the last one.

4.4.4 Technical Details

A professional cameraman from the University of Groningen recorded and edited the fragments. A Sony PXW-z150 4k XDCAM camera was used, the fragments were shot in full HD and with a resolution of 1920x1080 pixels.

Most of the situations were shot from the front, some of the items were recorded from a different angle. The fragments were supposed to contain as little distractions as possible, this resulted in no camera movement or combining two shots to make one fragment. The camera stayed on the same position for both literal and ironic condition, also during the whole fragment the camera did not move. In some items the camera had to zoom in or out a little bit to make the situations clearly visible.

The microphones used in the video fragments are EK 100 G3 camera receivers with adaptive diversity, SK 100 G3 body pack transmitters and ME 2 pin up microphones.

4.5 Audio Fragments

The questions were audio recorded to guarantee that every participant got the questions with the same intonation and tone, so there would be no focus due to the experimenter's pronunciation. Furthermore, since we tested children who are not (all) able to read, we could not use written questions.

For the audio fragments of the questions, a 24 year old female student was recorded while reading all the sentences aloud. Moreover, she spoke without an accent and with natural intonation. She was not aware of the research topic. We repeated this process three times, each with a random order of

sentences. We chose for each sentence the best recording based on, among others, audibility, natural intonation.

For the audio recordings of the questions the following equipment of the Groningen University were used: Shure SM27 capacitor microphone, Sonifex Redbox RB DA 6 amplifier and Sonifex S0 Broadcast mixer.

4.6 Procedure

The participants were tested individually in a quiet room. The children were tested at their school and they were told to play a computer game. The participants were instructed to answer the questions by saying 'YES' or 'NO', and that sometimes an additional open-ended question could be asked. As mentioned in Section 4.2.3, the questions were asked in particular order; *Q-intention*, *Q-meaning*, *(Q-justification)*, *Q-belief*, and *Q-emotion*.

Each participant got one of the four experiment lists. The participants received two practice items in random order before the experiment started. The target items occurred in random order. Each video and audio fragment was played once. During the experiment, the experimenter clicked on the buttons and wrote down notes. After the experiment the children may choose a sticker. The experiment took around 15 minutes.

5 Results

Every participant received 8 test items with a literal statement followed by four 'YES'-'NO' questions: *Q-intention*, *Q-meaning*, *Q-belief*, and *Q-emotion*. Furthermore, the participants received 8 test items with an ironic statement followed by the same four 'YES'-'NO' questions and the additional open-ended question *Q-justification*.

The questions were asked in the following order: *Q-intention*, *Q-meaning*, (*Q-justification*), *Q-belief*, and *Q-emotion*. However, we will present the results starting with the first-order ToM questions, giving the following order of results to be presented: *Q-meaning*, *Q-belief*, *Q-emotion*, *Q-intention*, and *Q-justification*.

We excluded 0.18% of the children's data (2 items * 4 questions = 8 observations) because at these items the participant did not see the video fragments due to interruptions. Due to similar interruptions, 1.17 % of the adults' data was removed (3 items * 4 questions = 12 observations).

The results of the 'YES'-'NO' questions were analyzed using generalized linear mixed-effect models (Bolker et al., 2009; Verzani, 2014; Faraway, 2016) in R (R Core Team, 2013). We used binomial family which leads to the estimates in logit scale where 0 is chance. The results of the open-ended question were analyzed using linear models.

In the results, two comparisons are made for each question. One between adults and children, with the adults as the intercept. The other between 6-, 7- and 8-year old children, with the 8-year olds as the intercept. With forward fitting followed by backward fitting the optimal model was found.

5.1 Meaning Question

In this section, the question *Q-meaning* is analyzed. *Q-meaning* is about the meaning of the statement: *'Does the woman mean [...]?'* As mentioned erlier, this is a first-order ToM question. The participant needs to understand that a speaker meant a ironic statement to be taken ironically.

5.1.1 Children versus Adults

Figure 5.1 reports the mean proportion of correct answers for children and adults per condition for *Q-meaning*. The mean proportion of correct answers of children (literal: 97.6%, ironic: 94.6%) is similar to the mean proportion of correct answers of adults (literal: 97.7%, ironic: 100%). This indicates that there is no large difference visible between the children and adults in both conditions.

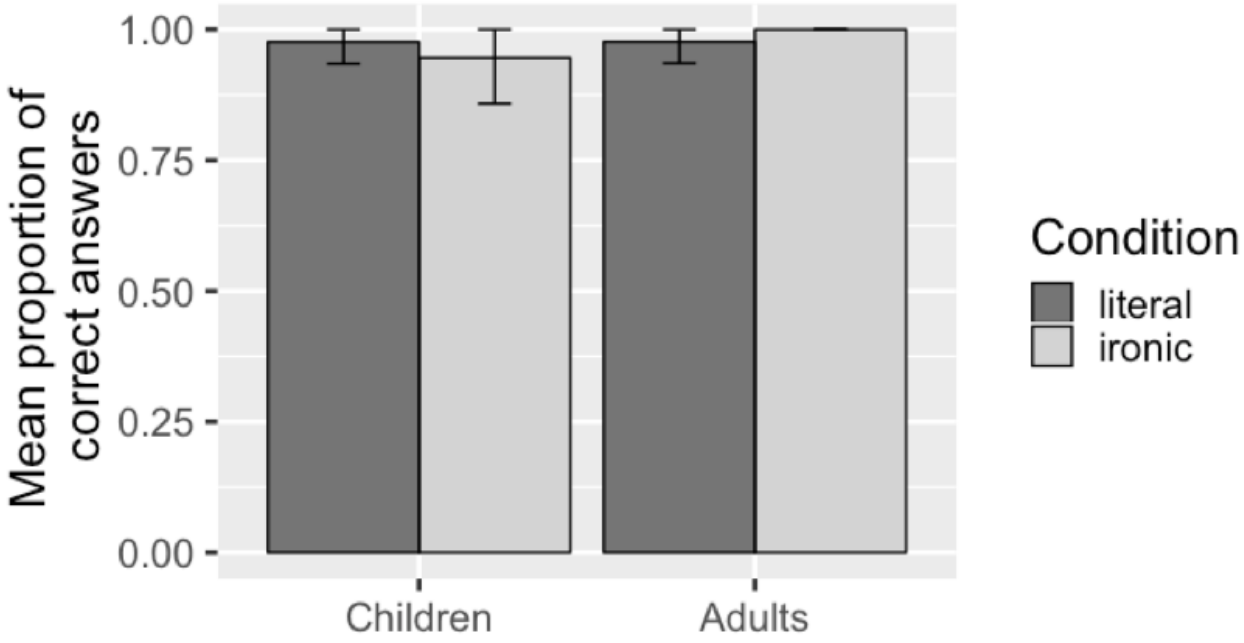

Figure 5.1: *Q-meaning*: Mean proportion of correct answers for children and adults for literal and ironic utterances (error bars indicating 1 Standard Deviation (SD)).

The model includes the predictors `condition` and `age`. Including the main effect `condition` explains significantly more variance ($\chi^2(1)=374.60$, p=.020) compared to the model without this main effect. Furthermore, including the main effect `age` explains significantly more variance ($\chi^2(1)=374.60$, p=.018) than the model without this main effect. Besides the two main effects, it also includes a random effect for the slope of the `participants:condition`.

Table 5.1 reports the results of the model, in which the intercept is the literal condition for adults. The intercept is significantly different from 0 (z=3.328, p<.001). The ironic condition does not significantly differ from the intercept (z=-1.750, p=.080). Furthermore children do not significantly differ from adults (z=-1.729, p=.084).

The results indicate that adults correctly answer the *Q-meaning* in both the literal and ironic condition. Moreover, the children do not differ from adults in the literal nor in the ironic conditions. Hence, these results show that children and adults have a high accuracy on *Q-meaning* in both the literal and ironic condition. Additionally, there is no difference between children and adults.

Table 5.1: *Q-meaning*: Fixed and random effects of the model for accuracy. The reference variable is 'literal condition for adults'.

Question: *Q-meaning*					
Data: Children and adults					
Model: $accuracy \sim condition + age + (0 + condition \mid participants)$					
Predictor		**Estimate**	**Std. Error**	**z value**	$Pr(>\|z\|)$
Intercept	***	9.083	2.729	3.328	0.000874
Condition (Ironic)		-3.446	1.969	-1.750	0.080107
Age (Children)		-2.196	1.270	-1.729	0.083792
.<.1, *p<.05, **p<.01, ***p<.001					

5.1.2 6-y.o. versus 7-y.o. versus 8-y.o.

Figure 5.2 reports the mean proportion of correct answers for 6-, 7-, and 8-years old children per condition for *Q-meaning*. It can be seen from this figure that 7-year olds (literal: 99.0%, ironic: 95.2%) have similar performance to 8-year olds (literal: 97.5%, ironic: 96.1%). However, 6-year olds have slightly lower mean proportions of correct answers (literal: 95.3%, ironic: 88.6%).

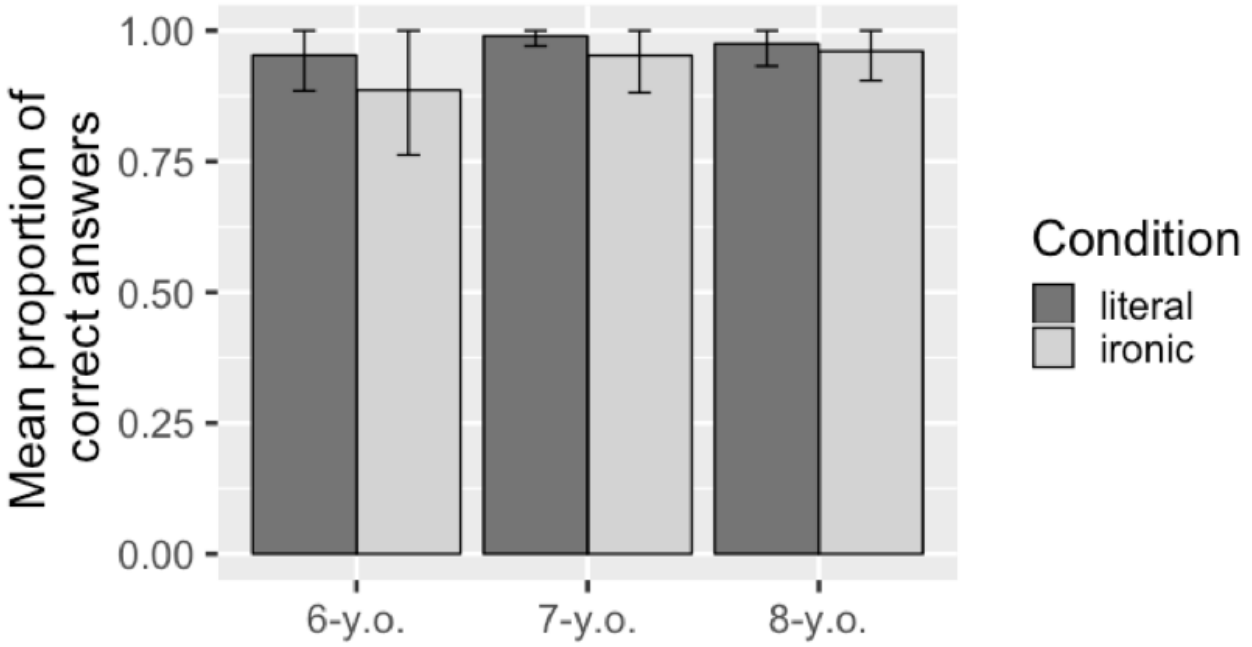

Figure 5.2: *Q-meaning*: Mean proportion of correct answers for 6-, 7-, and 8-y.o. children for literal and ironic utterances (error bars indicating 1 SD).

The model includes the predictors `condition`, `age`, and a random intercept for `participants`. Including the main effect `condition` explains significantly more variance ($\chi^2(1)$=346.79, p=.006)

than the model without this main effect. Including the main effect `age` is not significantly different than the model without this main effect. Table 5.2 shows that the main effect `age` is not significant.

Table 5.2 provides the summary of the model with the literal condition for 8-year olds as intercept (z=10.097, p<.001), which is significantly different from 0. The ironic statement condition is significantly lower than the intercept (z=-2.603, p<.01).

The results indicate that 8-year olds correctly answer *Q-meaning* in both the literal and ironic condition. Additionally it shows that 6- and 7-year olds are not different from 8-year olds in performance on *Q-meaning*. However, children perform slightly worse when the statement is ironic.

Table 5.2: *Q-meaning*: Fixed and random effects of the model for accuracy. The reference variable is 'literal condition for 8-year olds'.

Question: *Q-meaning* Data: Children Model: $accuracy \sim condition + age + (1\|participants)$					
Predictor		**Estimate**	**Std. Error**	**z value**	$Pr(>\|z\|)$
Intercept	***	4.2812	0.4240	10.097	< 2e-16
Condition (Ironic)	**	-0.8973	0.3447	-2.603	0.00925
Age (6-y.o.)	.	-0.9202	0.5146	-1.788	0.07377
Age (7-y.o.)		0.1315	0.4761	0.276	0.78242
.<.1, *p<.05, **p<.01, ***p<.001					

5.2 Belief Question

In the current section, *Q-belief* is analyzed: *'Does the woman think [...]?'*. The participant needs to understand the first-order belief of the speaker to correctly answer this question.

5.2.1 Children versus Adults

Figure 5.3 reports the mean proportion of correct answers for children and adults per condition for *Q-belief*. In this graph there is no clear difference between the mean proportion of correct answers of children (literal: 88.8%, ironic: 88.5%) and adults (literal: 89.1%, ironic: 92.1%). Moreover, it can be seen that both groups do not perform at ceiling.

The model includes the predictors `condition` and `age`. Including these two main effects is not significantly different from the intercept model. The summary of the model is represented in Table 5.3 and it shows that the main effects `condition` and `age` are not significant. The model also includes random intercepts for `participants` and `items`. The intercept is significantly different from 0 (z=6.251, p<.001).

The results indicate that adults correctly answer *Q-belief* in both the literal and ironic condition. Furthermore, the children do not differ from adults in the literal nor the ironic conditions. Hence, these results show that children and adults have a high accuracy in the both literal and ironic condition for *Q-belief*. Furthermore there are no differences between children and adults.

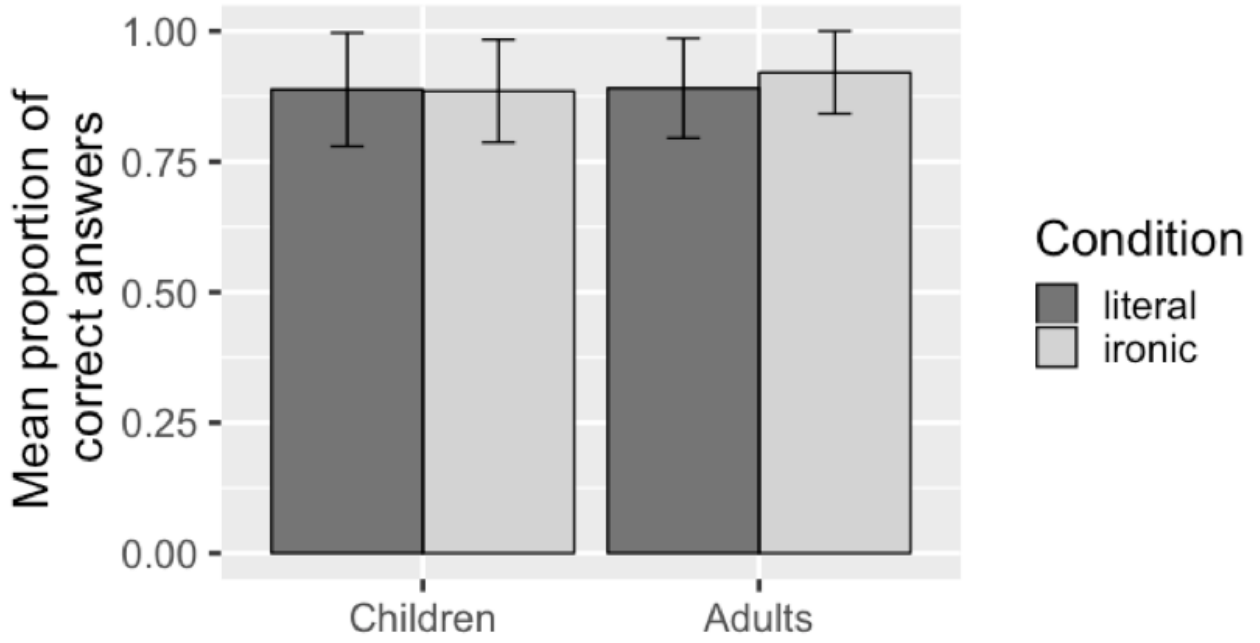

Figure 5.3: *Q-belief*: Mean proportion of correct answers for children and adults for literal and ironic utterances (error bars indicating 1 SD).

Table 5.3: *Q-belief*: Fixed and random effects of the model for accuracy. The reference variable is 'literal condition for adults'.

Question: *Q-belief*					
Data: Children and adults					
Model: $accuracy \sim condition + age + (1|participants) + (1|items)$					
Predictor		**Estimate**	**Std. Error**	**z value**	$Pr(>\|z\|)$
Intercept	***	3.08443	0.49346	6.251	4.09e-10
Condition (Ironic)		0.04351	0.18343	0.237	0.813
Age (Children)		-0.24516	0.33451	-0.733	0.464
.<.1, *p<.05, **p<.01, ***p<.001					

5.2.2 6-y.o. versus 7-y.o. versus 8-y.o.

Figure 5.4 reports the mean proportion of correct answers for 6-, 7-, and 8-year old children per condition for *Q-belief*. In this graph, there is again no obvious difference between the results of the different age groups and the conditions (respectively 6-, 7- and 8-year olds for literal items: 84.6%, 89.1%, 90.0% and ironic items: 90.9%, 86.8%, 88.9%).

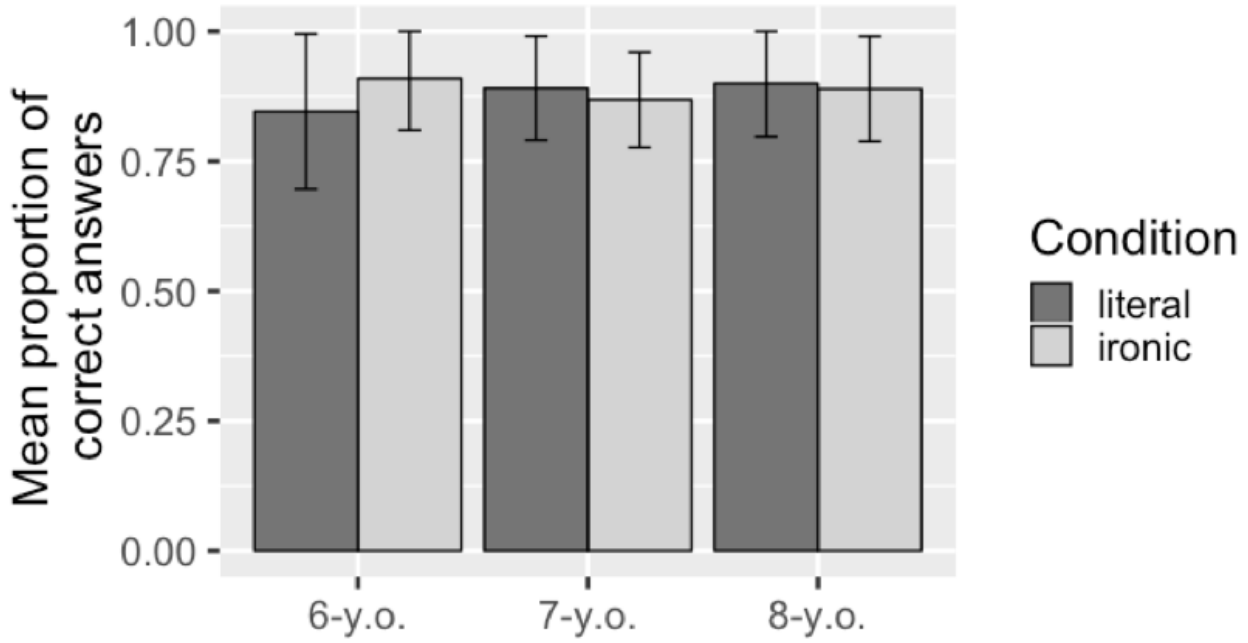

Figure 5.4: *Q-belief*: Mean proportion of correct answers for 6-, 7-, and 8-y.o. children for literal and ironic utterances (error bars indicating 1 SD).

Table 5.4: *Q-belief*: Fixed and random effects of the model for accuracy. The reference variable is 'literal condition for 8-year olds'.

Question: *Q-belief* Data: Children Model: $accuracy \sim condition + age + (1\|participants) + (1\|items)$					
Data: Children					
Predictor		**Estimate**	**Std. Error**	**z value**	$Pr(>\|z\|)$
Intercept	***	2.87413	0.41119	6.990	2.75e-12
Condition (Ironic)		-0.04421	0.19916	-0.222	0.824
Age (6-y.o.)		-0.12120	0.39761	-0.305	0.760
Age (7-y.o.)		-0.19253	0.30238	-0.637	0.524
.<.1, *p<.05, **p<.01, ***p<.001					

The model includes the predictors `condition` and `age`. Including these two main effects is not significantly different from the intercept model. The summary of the model is represented in Table 5.4 and it shows that the main effects `condition` and `age` are not significant. The model also includes random intercepts for `participants` and `items`. The intercept is significantly different from 0 (z=6.990, p<.001).

The results indicate that 8-year olds correctly answer *Q-belief* in both the literal and ironic condition. Furthermore, the 6- and 7-year old children do not differ from 8-year olds in the literal nor

the ironic conditions. Hence, these results show that 6-, 7-, and 8-year olds have a high accuracy in both the literal and ironic condition for *Q-belief*. Furthermore, there is no difference in performance on *Q-belief* between 6-, 7-, and 8-year old children.

5.3 Emotion Question

In this section, question *Q-emotion* is discussed: *'Is the woman grumpy/happy/irritated?'*. To correctly answer this question, the participant needs to understand the emotion of the speaker.

5.3.1 Children versus Adults

The mean proportion of correct answers for children and adults are shown in Figure 5.5. This graph does not show any differences between the children's mean proportion of correct answers (literal: 88.9%, ironic: 88.5%) and adults' mean proportion of correct answers (literal: 84.4%, ironic: 89.7%). Moreover, both groups do not perform at ceiling.

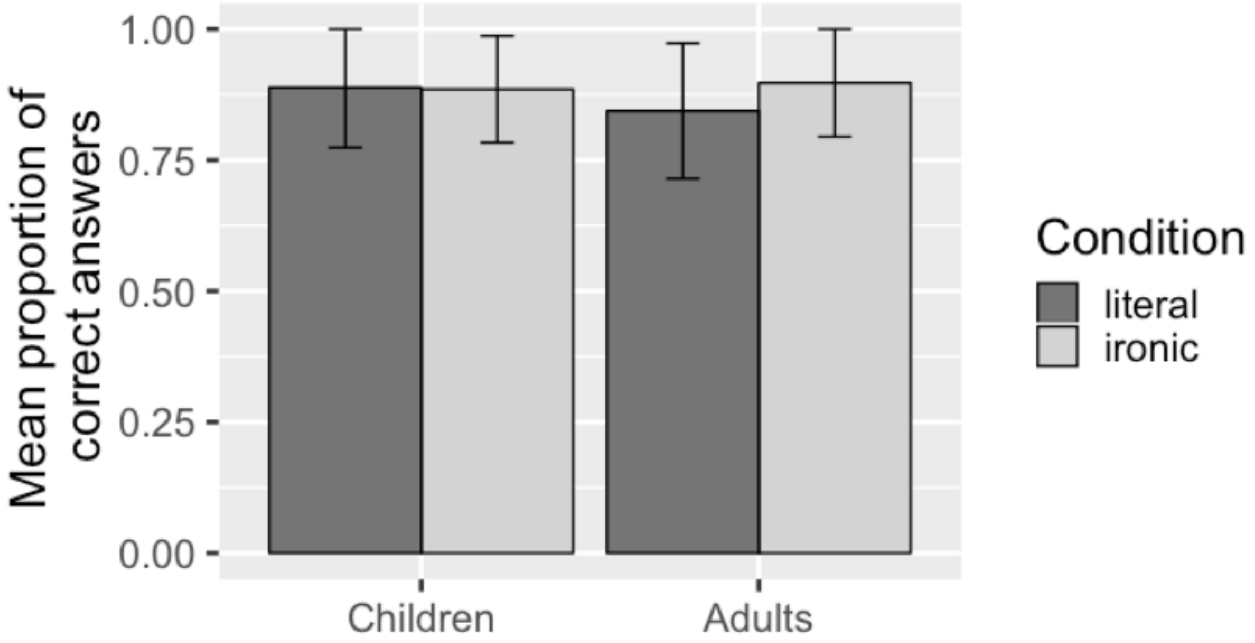

Figure 5.5: *Q-emotion*: Mean proportion of correct answers for children and adults for literal and ironic utterances (error bars indicating 1 SD).

The model includes the predictors `condition` and `age`. Including these two main effects is not significantly different from the intercept model. The summary of the model is represented in Table 5.5 and it shows that the main effects `condition` and `age` are not significant. The model also includes random intercepts for `participants` and `items`. The intercept is significantly different from 0 ($z=5.455$, $p<.001$).

The results indicate that adults correctly answer *Q-emotion* in both the literal and ironic condition. Furthermore, the children do not differ from adults in the literal nor the ironic conditions. Hence, these results show that children and adults have a high accuracy in the literal and ironic condition for *Q-emotion*. Furthermore there are no differences between children and adults.

Table 5.5: *Q-emotion*: Fixed and random effects of the model for accuracy. The reference variable is 'literal condition for adults'.

Question: *Q-emotion* Data: Children and adults Model: $accuracy \sim condition + age + (1\|participants) + (1\|items)$					
Data: Children and adults					
Predictor		**Estimate**	**Std. Error**	**z value**	$Pr(>\|z\|)$
Intercept	***	2.83060	0.51892	5.455	4.9e-08
Condition (Ironic)		0.02182	0.18401	0.119	0.906
Age (Children)		0.17313	0.34176	0.507	0.612
.<.1, *p<.05, **p<.01, ***p<.001					

5.3.2 6-y.o. versus 7-y.o. versus 8-y.o.

Figure 5.6 reports the mean proportion of correct answers for 6-, 7-, and 8-years old children per condition for *Q-emotion*. Again, this graph does not show obvious differences between the conditions (respectively 6-, 7-, 8-year olds in literal items: 88.1% 88.5%, 89.3% and ironic items: 83.0%, 84.2%, 93.2%).

The model includes the predictors `condition` and `age`. Including these two main effects is not significantly different from the intercept model. The summary of the model is represented in Table 5.6 and it shows that the main effects `condition` and `age` are not significant. The model also includes random intercepts for `participants` and `items`. The intercept is significantly different from 0 (z=7.380, p<.001).

The results indicate that 8-year olds correctly answer *Q-emotion* in the literal and ironic condition. Furthermore, the 6- and 7-year old children do not differ from 8-year olds in the literal nor the ironic conditions. Hence, these results show that 6-, 7-, and 8-year old children have a high accuracy in the literal and ironic condition for *Q-emotion*. Furthermore there are no differences between 6-, 7-, and 8-year olds.

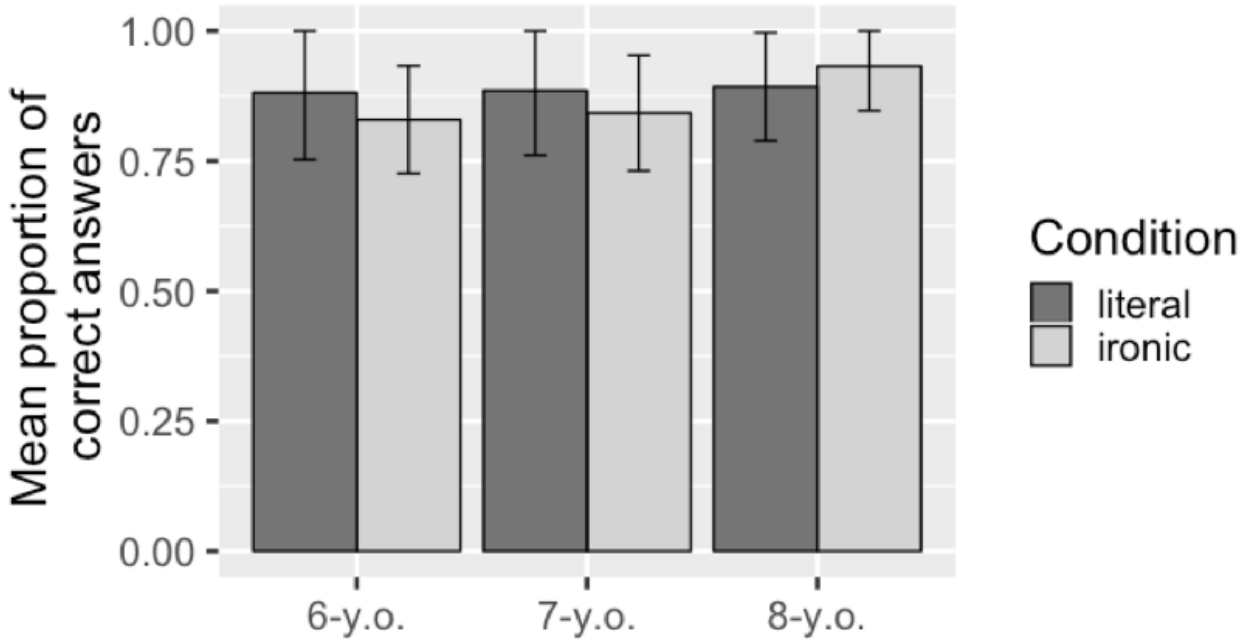

Figure 5.6: *Q-emotion*: Mean proportion of correct answers for 6-, 7-, and 8-y.o. children for literal and ironic utterances (error bars indicating 1 SD).

Table 5.6: *Q-emotion*: Fixed and random effects of the model for accuracy. The reference variable is 'literal condition for 8-year olds'.

Question: *Q-emotion*					
Data: Children					
Model: $accuracy \sim condition + age + (1|participants) + (1|items)$					
Predictor		**Estimate**	**Std. Error**	**z value**	$Pr(>\|z\|)$
Intercept	***	3.21234	0.43528	7.380	1.58e-13
Condition (Ironic)		-0.09117	0.20253	-0.450	0.6526
Age (6-y.o.)	.	-0.65904	0.38847	-1.697	0.0898
Age (7-y.o.)	*	-0.61785	0.30362	-2.035	0.0419
.<.1, *p<.05, **p<.01, ***p<.001					

5.4 Intention Question

In this section, the *Q-intention* is analyzed. This question is a question about second-order intention of the speaker: *'The woman says to the man:* `[statement]` *Does she want the man to believe this?* To answer this question correctly, the participant needs to be able to reason second-order Theory of Mind.

5.4.1 Children versus Adults

Figure 5.7 shows the mean proportion of correct answers for children and adults per condition for *Q-intention*. Descriptive statistics show that the adults have a high mean proportion of correct answers in both conditions (literal: 95.3%, ironic: 94.5%). The children's mean proportion of correct answers in the literal condition is also high (92.2%). However, children's mean proportion of correct answers in the ironic condition is lower (36.0%) compared to the literal condition.

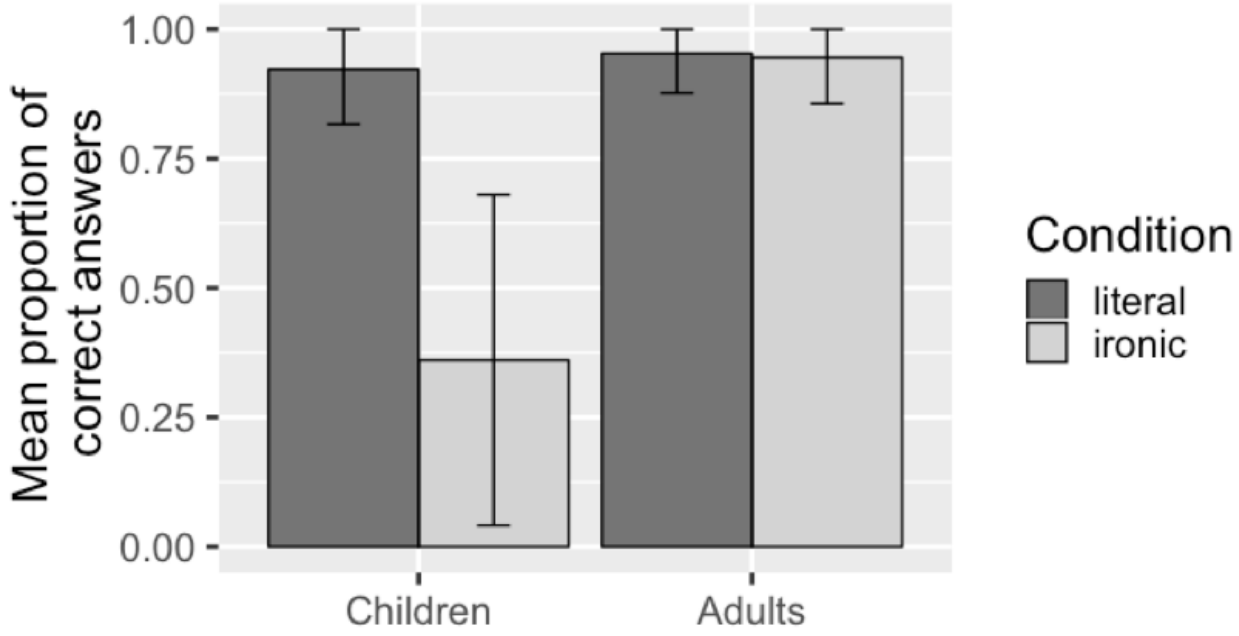

Figure 5.7: *Q-intention*: Mean proportion of correct answers for children and adults for literal and ironic utterances, (error bars indicating 1 SD).

Besides a lower mean proportion of correct answers for children in the ironic condition, it also stands out in Figure 5.7 is that the error bars for the ironic condition for children is much larger than the other error bars in this figure. The error bars are calculated by the standard deviation (SD). The large standard deviation in the children's ironic condition means that in this case the proportion of correct answers vary more from the average than in the other conditions. In other words, the data of the children's ironic condition is dispersed over a wide range of values.

Figure 5.7 represents the means in each condition. Therefore from this graph we can not see if there is a binomial distribution of the data in this graph. Given the wide SD in the children's irony condition we investigated its cause by looking further at the answers of individual children in order to determine if it was caused by variation within children or between children. We checked the number of children per number of correct answers (see Figure 5.8). It is apparent from this figure that the children are neither always correct nor always incorrect at this question.

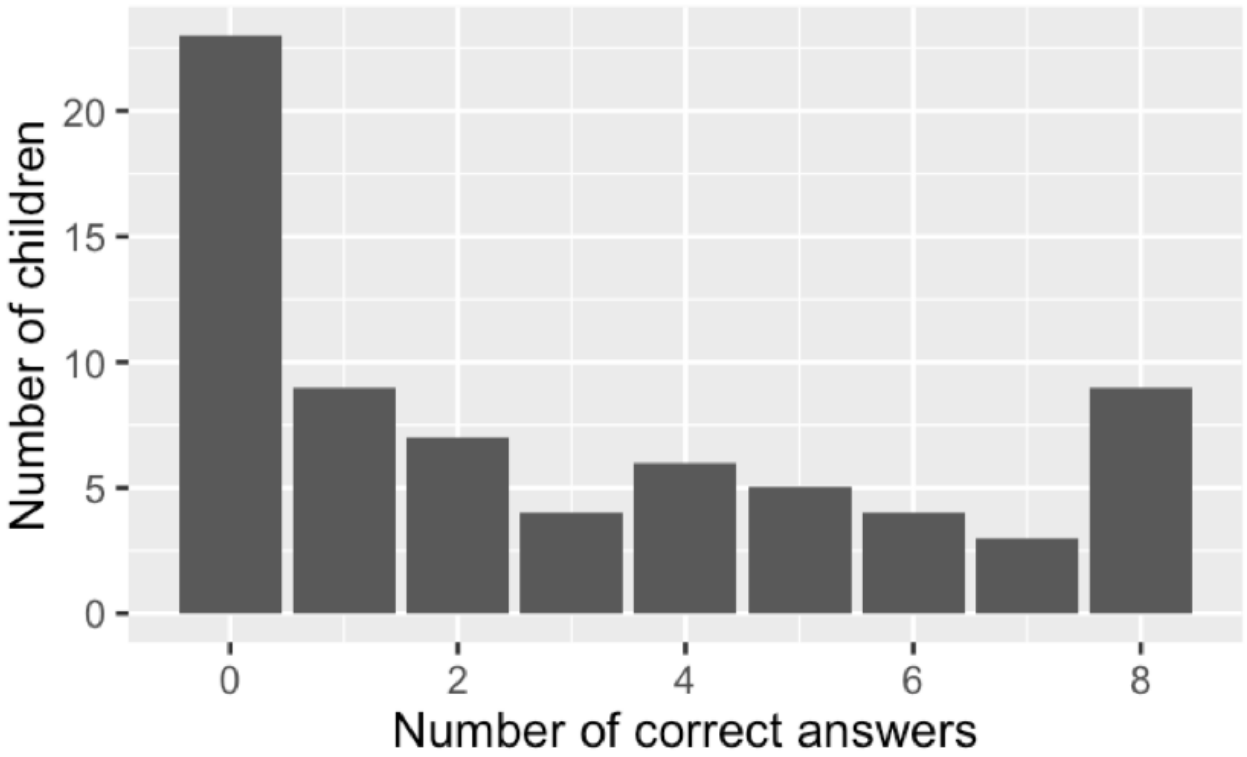

Figure 5.8: *Q-intention*: The number of children ($N = 70$) per number of correct answers (0-8) in ironic condition. One child did not give an answer for one item, so (s)he contributed only 7 responses.

Another reason for a mean proportion of 36% for children's responses on ironic items (Figure 5.7) could be that some items are much more difficult. This would result in fewer correct answers for that item. Figure 5.9 presents the percentage of correct answer per item. Here we see that there is some variation in the percentage of correct answers between the items. However, this difference is not extraordinary, so the low mean proportion of correct answers of children in the ironic statement condition is also not explained by particular items being difficult.

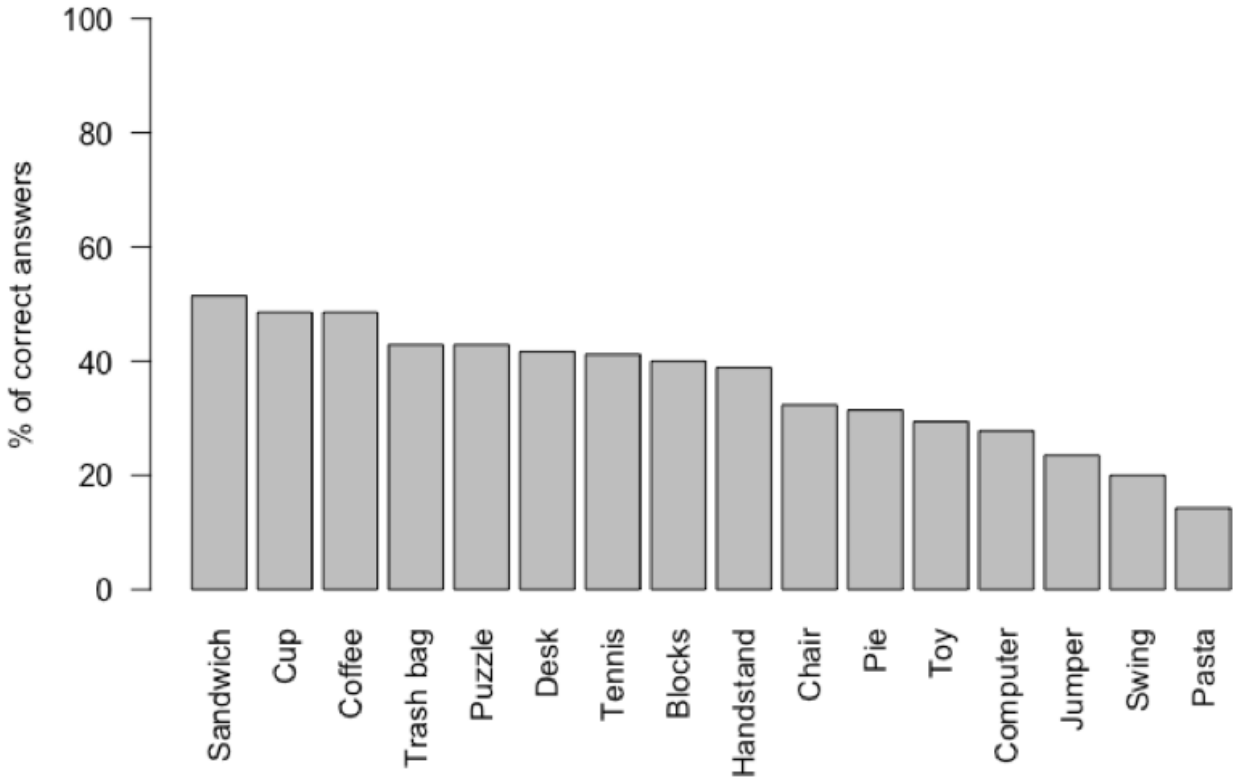

Figure 5.9: *Q-intention*: Percentage of correct answers per item in the ironic condition. Every item had around 35 answers.

The model includes the predictors `condition` and `age`. Including the interaction between `condition` and `age` explains significantly more variance ($\chi^2(1)=865.84$, $p<.001$) than the model without the interaction. Furthermore, it includes a random effect for the slope of the `participants:condition`.

Table 5.7 reports the results of the model. The intercept is the literal condition for adults, which is significantly different from 0 (z=5.003, $p<.001$). The ironic condition for adults does not significantly differ from the intercept (z=.362, p=.717), neither does the literal condition for children (z=-0.864, p=.388). However, the mean proportion of correct answers in the ironic condition for children is significantly lower than the intercept (z=-3.631, $p<.001$), it is even below chance.

The results of *Q-intention* indicate that adults correctly answer *Q-intention* in both the literal and ironic condition. Furthermore, the results also show that children do not perform differently from adults in the literal condition. However, for ironic items children do perform significantly lower than adults in terms of how well they recognize that the target sentence was non-literal. Hence, children's accuracy for *Q-intention* in literal items is similar to adults. However, children's accuracy in ironic items is much lower than adult's accuracy for ironic items.

Table 5.7: *Q-intention*: Fixed and random effects of the model for accuracy. The reference variable is 'literal condition for adults'.

Question: *Q-intention*					
Data: Children and adults					
Model: $accuracy \sim condition * age + (0 + condition \mid participants)$					
Predictor		**Estimate**	**Std. Error**	**z value**	$Pr(>\|z\|)$
Intercept	***	4.6867	0.9313	5.033	4.84e-07
Condition (Ironic)		0.5388	1.4869	0.362	0.717096
Age (Children)		-0.7708	0.8926	-0.864	0.387818
Condition (Ironic) : Age(Children)	***	-5.7141	1.5736	-3.631	0.000282
.<.1, *p<.05, **p<.01, ***p<.001					

5.4.2 6-y.o. versus 7-y.o. versus 8-y.o.

Figure 5.10 reports the mean proportion of correct answers for 6-, 7-, and 8-years old children per condition for *Q-belief*. In this graph, no obvious difference between the 6-, 7- and 8-year olds can be seen. However, the graph shows that in each age group the mean proportion of correct answers is lower in the ironic condition than in the literal condition (respectively 6-, 7- and 8-year old children in literal items: 85.9%, 92.7%, 93.9% and in ironic items: 26.1%, 37.4%, 38.2%).

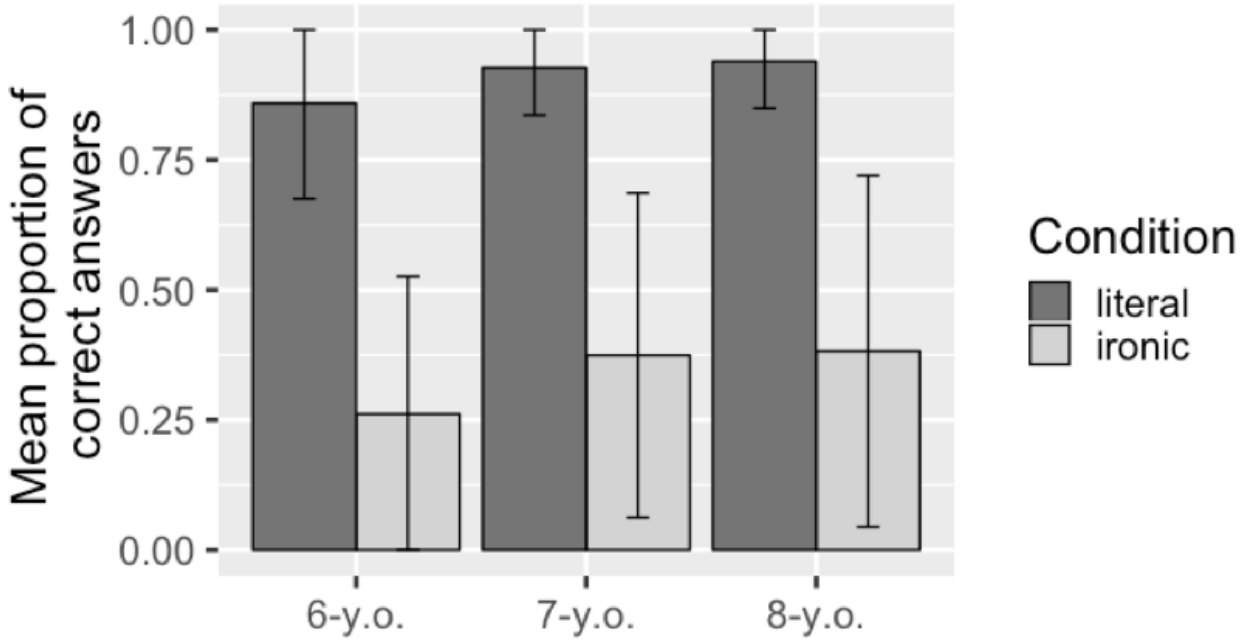

Figure 5.10: *Q-intention*: Mean proportion of correct answers for 6-, 7-, and 8-y.o. children for literal and ironic utterances (error bars indicating 1 SD).

The model includes the main effects `condition` and `age`. It includes also a random slope for `participants:condition`. Including the main effect `condition` explains significantly more variance than the model without the main effect (χ^2(1)=781.02, p<.001). Including the main effect `age` is not significantly different than the model without this main effect. Table 5.2 shows that the main effect `age` is not significant.

Table 5.8 shows the summary of this model. The intercept is the literal condition for 8-year olds, which is significantly different from 0 (z=7.352, p<.001). The accuracy in the ironic statement condition is significantly lower than the intercept (z=-7.230, p<.001), it is even below chance.

The results of *Q-intention* indicate that 8-year old children correctly answer *Q-intention* in the literal condition. However, they have a much lower accuracy in the ironic condition. This pattern is also seen with 6- and 7-year olds. Hence, 6-, 7-, and 8-year olds have a lower accuracy for *Q-intention* in ironic items than in literal items, but there is no difference per age group.

Table 5.8: *Q-intention*: Fixed and random effects of the model for accuracy. The reference variable is 'literal condition for 8-year olds'.

Question: *Q-intention* Data: Children Model: $accuracy \sim condition + age + (0 + condition \mid participants)$					
Predictor		**Estimate**	**Std. Error**	**z value**	$Pr(>\|z\|)$
Intercept	***	4.0437	0.5500	7.352	1.95e-13
Condition (Ironic)	***	-5.0374	0.6967	-7.230	4.82e-13
Age (6-y.o.)	.	-0.9933	0.5778	-1.719	0.0856
Age (7-y.o.)		-0.2972	0.4418	-0.673	0.5011
.<.1, *p<.05, **p<.01, ***p<.001					

5.5 Justification Question

In this section, the open-ended justification for speaker's non-literal statement question is discussed. *Q-justification* is: *'Why does (s)he say this?'*. Because this question is an open-ended question, the answers were categorized. For this question the participant needs to understand the reason why the speaker utters a non-literal statement. Table 5.9 represents the category with score and the type of justification (this table is the same as in Section 4.2.4).

The justification for the speaker's non-literal statement was only asked in the ironic statement condition. Only adults, 8-year olds and twenty-two 7-year olds were asked *Q-justification*. Due to human error, such as playing the next question before asking *Q-justification*, not every participant had answered this question for every ironic item. This leads to an unequal amount of justification observations per participants, resulting in 439 (84.42%) observations.

Table 5.9: Each category with their score and the type of justification.

Category	**Score**	**Type of justification**
1a) No Answer 1b) Other 1c) Literal	0	- The participant does not give an answer to the question. - The participants gives an answer which is not categorizable. - The participant says that the speaker tells the truth.
2a) Intonation 2b) Proverb 2c) Humor	1	The participant knows that the statement should not taken literally, however, the participant does not say anything about the reason.
3) Opposite	2	The participant knows that the speaker means the opposite but does not say anything about the intention.
4a) Politeness 4b) White Lie	3	The participant knows that the speaker means the opposite, however, the participants does think that the intention is something else than ironic.
5) Ironic	4	The participants knows that the speaker means the opposite and that the intention of the speaker is to let the listener know (s)he means the opposite.

5.5.1 Children versus Adults

Figure 5.11 shows the percentage of answers per category for adults and children. Adults mostly gave answers categorized as Ironic (88.98%). Children mostly gave answers categorized as Opposite (40.44%), No Answer (15.57%), and Humor (12.84%).

To analyze the results of *Q-justification*, scores were given to each answer based on the type of justification for the response on the second-order intention, with lower scores given for more literal responses and higher scores given for answers that showed a clearer understanding of the irony (see Table 5.9). For every participant, an average score was calculated.

In Figure 5.12 the mean scores for children and adults are shown. In this graph, there is a difference in mean score between children (1.70) and adults (3.82). It can be seen from the error bars that the children's scores are more spread from the mean than the adult's scores. To check if the difference between children and adults is significant, we analyzed the average score per participant.

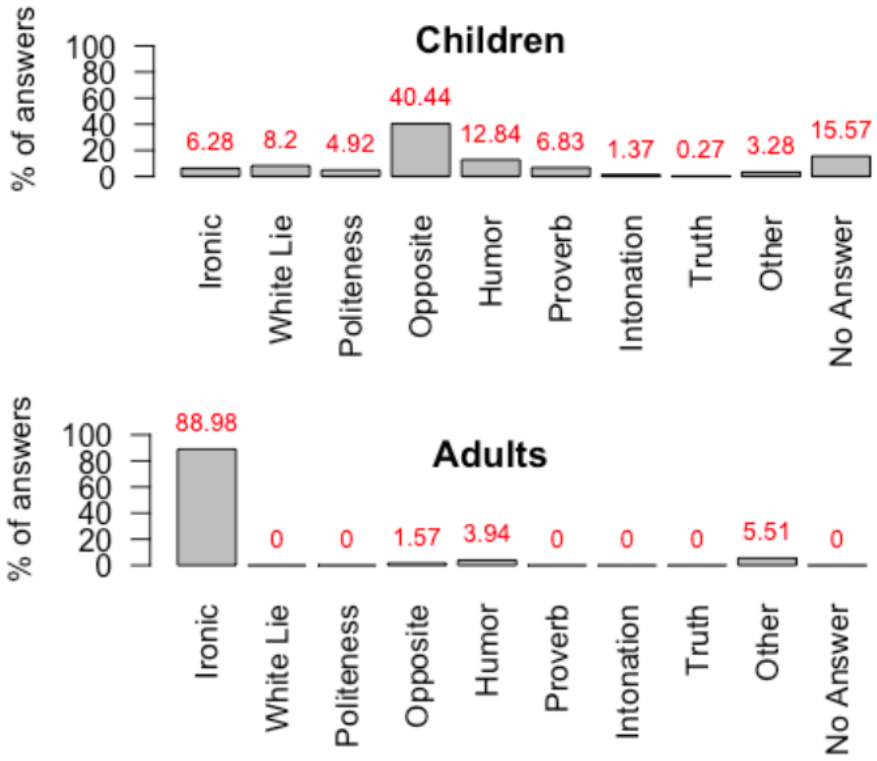

Figure 5.11: Percentage of answers per category for children (top) and adults (bottom). Number of observations: children=366 and adults=127.

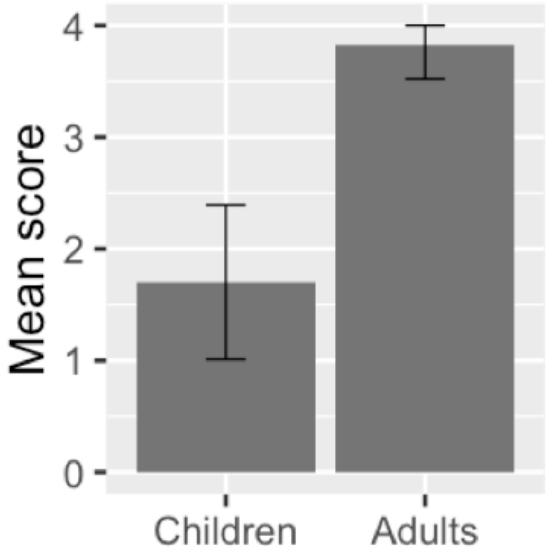

Figure 5.12: *Q-justification*: For every participant, an average score was calculated. This graph shows the mean scores for children and adults (error bars indicating 1 SD). The scores are given based on the type of justification for the response on the second-order intention question (Table 5.9).

The results were analyzed using linear models. The model includes the predictor `age` ($\chi^2(1)$=55.921, p<.001). Table 5.10 shows the summary of the model with adults as intercept (t=18.54, p<.001), which is significantly different from 0. The children's average score is significantly lower than the intercept (t=-9.08, p<.001).

The results of *Q-justification* indicate that adults have a high mean score and that adult's justification for speaker's non-literal statement is that the speaker was being ironic. Furthermore, the results also show that children perform differently from adults. They have a lower mean score, and their justifications are different from adults.

Table 5.10: *Q-justification*: Fixed effect of the model for accuracy. The reference variable is 'adults'.

Question: *Q-justification* Data: Children and adults Model: *averageScore* ~ *age*					
Predictor		**Estimate**	**Std. Error**	**t value**	$Pr(>\|t\|)$
Intercept	***	3.8177	0.2059	18.54	< 2e-16
Age (Children)	***	-2.1157	0.2330	-9.08	1.67e-13
.<.1, *p<.05, **p<.01, ***p<.001					

5.5.2 7-y.o. versus 8-y.o.

Figure 5.13 shows the percentage per category for 7- and 8-year olds. It appears that the 7-year olds are mostly answering with Opposite (38.67%), followed by Humor (20.67%), No Answer (19.33%), and White Lie (12.67%). Eight-year olds often gave answers which fall in the category Opposite (41.67%). Furthermore, they also answer with Inappropriate (12.96%). In contrast with 7-year olds, 8-year old children identify the ironic statement as Ironic in 10.65% of the cases.

In Figure 5.14 the mean scores for 7- and 8-year olds. The mean score of 7-year olds (1.51) and 8-year olds (1.82) do not seem to differ much. In addition, the SD's are similar, which can be seen by the comparable size of the error bars.

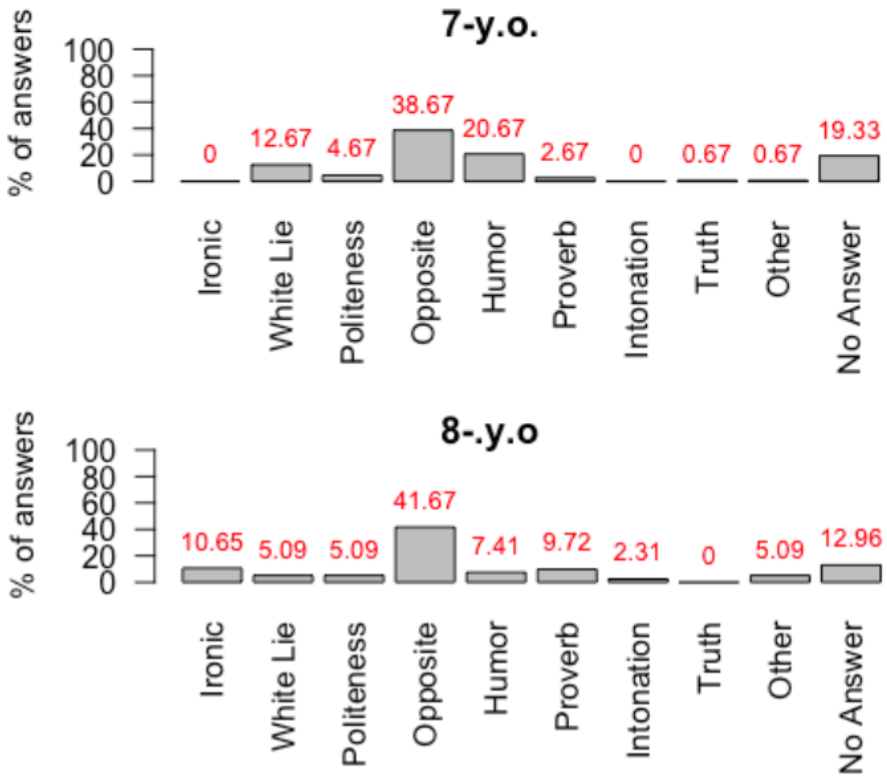

Figure 5.13: Percentage of answers per category for 7-y.o. (top) and 8-y.o. (bottom). Number of observations: 7-y.o.=150 and 8-y.o.=216.

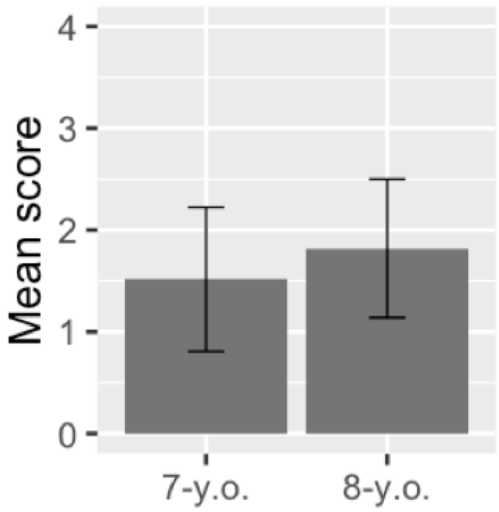

Figure 5.14: *Q-justification*: For every participant, an average score was calculated. This graph shows the mean scores for 7-y.o. and 8-y.o. (error bars indicating 1 SD). The scores are given based on the type of justification for the response on the second-order intention question(Table 5.9).

The model (Table 5.11) includes the main effect `age`. Including this main effects is not significantly different from the intercept model. Table 5.11 shows that the main effect `condition` is not significant. The intercept is significantly different from 0 (t=14.27, p<.001).

While the difference in mean score per person between 7- and 8-year olds are not significant (Figure 5.14), the difference in the categories are visible in Figure 5.13. Hence, the results of *Q-justification* show that 7- and 8-year old children have similar mean score. However there is a qualitative difference in the type of answers they give (Figure 5.13).

Table 5.11: *Q-justification*: Fixed effect of the model for accuracy. The reference variable is '8-year olds'.

Question: *Q-justification* Data: Children Model: $averageScore \sim age$					
Predictor		**Estimate**	**Std. Error**	**t value**	$Pr(>\|t\|)$
Intercept	***	1.8193	0.1515	12.008	< 2e-16
Age (7-y.o.)		-0.3039	0.2439	-1.246	0.218
.<.1, *p<.05, **p<.01, ***p<.001					

5.6 Relation between Intention and Justification

Children's responses to the *Q-intention* and *Q-justification* could be related. *Q-intention* is about whether the speaker want the listener to believe the literal meaning of the statement, *Q-justification* is about the reason of the speaker to utter a non-literal statement. To see whether there is evidence of a quantitative relation we look for a correlation between participant's accuracy on *Q-intention* and his/her score on *Q-justification* (see Figure 5.15).

One surprising aspect of Figure 5.15 is that there is one participant who had a mean accuracy of 0 on *Q1-intention*, yet, this participant had a mean score of 4 on *Q-justification*.

From this graph, a clear relation between *Q-intention* and *Q-justification* could not be identified. We did not statistically test this theory.

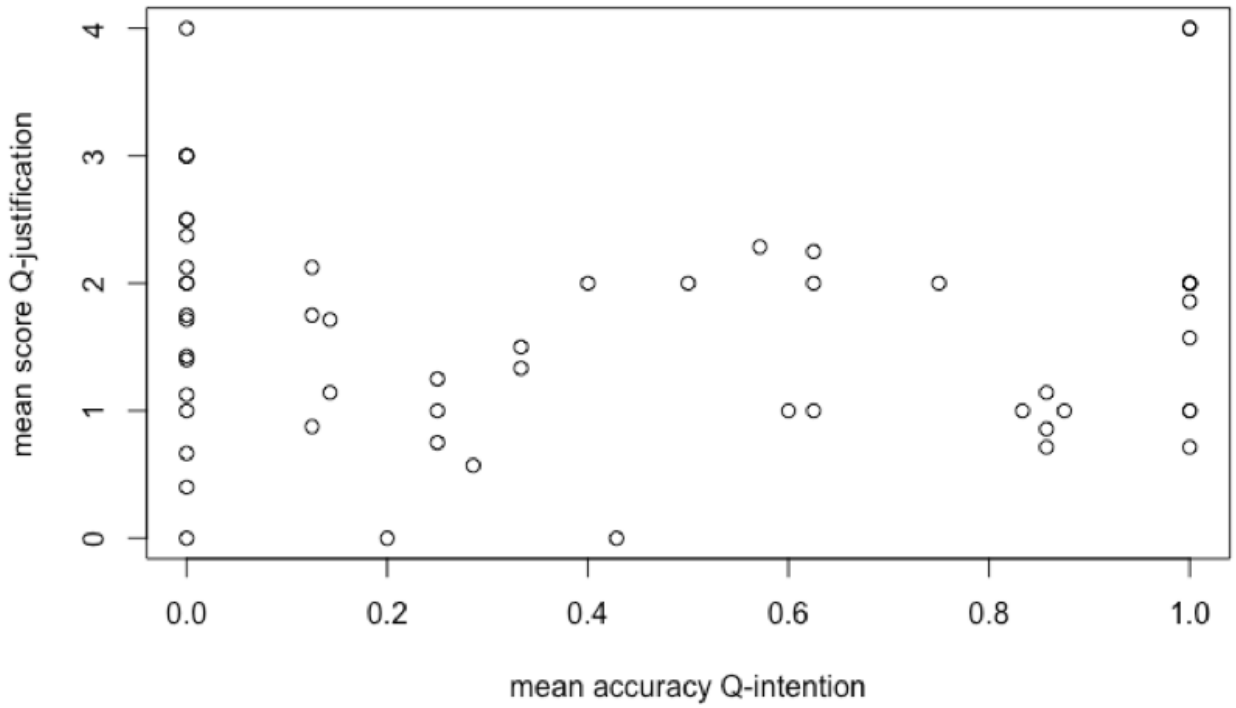

Figure 5.15: Children's mean accuracy on *Q-intention* and their mean score on *Q-justification* in the ironic condition.

6 Discussion

6.1 Findings

The aim of the present research was to examine how children will perform on an irony task when the experiment requires less cognitive resources than previous studies. The second aim of this study was to determine children's development in irony understanding. Additionally, we were also interested in children's justifications for the speaker's non-literal statement. To address these aims we formulated two research questions.

- *Research Question 1:* Can 6- to 8-year old children show adult-like irony understanding?
- *Research Question 2:* Do 6- to 8-year old children develop with age in irony understanding?

Five components of irony understanding were identified and the two research questions contained both five sub-questions with respect to these components. The components are the following:

a) Understanding that the meaning of an ironic statement has to be taken non-literal.

b) Judging first-order belief of an ironic speaker.

c) Identifying emotions expressed by an ironic speaker.

d) Understanding the second-order intention of an ironic speaker.

e) Understanding the justification for speaker's non-literal statement.

The first research question was whether children show adult-like irony understanding in our experiment. In our hypothesis, we expected that children are adult-like at components a), b), and c) but are not adult-like at components d) and e).

The results confirm our hypothesis. The results show that children are adult-like in recognizing a non-literal statement, judging the first-order belief of an ironic speaker, and identifying emotions expressed by an ironic speaker. However, children are not adult-like in understanding the second-order intention of an ironic speaker, which is in line with previous findings (Ackerman, 1983; Hancock et al., 2000; Filippova & Astington, 2008). Furthermore, children are not adult-like in their justifications for speaker's non-literal statements. Children's justifications for speaker's non-literal statement were mainly that the speaker meant the opposite (40.44%) or made a joke (12.84%). Adults' justifications were that the speaker was being ironic (89.98%). The opposite and a joke seem like a good description of irony. However, children responses do not indicate that they understand the second-order intention of an ironic speaker. The second-order intention of the speaker is an important difference between a joke and an ironic statement.

In other words, although 6- to 8-year old children are able to recognize that an ironic statement should not be taken literally, they do not understand the intention of a speaker and do not give a correct justification for why the speaker utters a non-literal statement. The fact that children do not take an ironic statement literally is also reflected in their justifications for a speaker's non-literal statement, only one time a child said that the speaker meant what (s)he said.

Another finding that can be seen is that despite children are adult-like in identifying the emotion of an ironic speaker, they do not understand the second-order intention of an ironic speaker. Moreover, even if children understand the intention of an ironic speaker, this does not mean that they also give a correct justification for speaker's non-literal statement. Hence, these results indicate that children are not adult-like at irony understanding.

The second research question of this study was whether 6- to 8-year old children show development in irony understanding. We expected that older children are more adult-like than younger children in irony understanding. Our results show that there are no quantitative differences in the five components between younger and older children. This means that the development of irony understanding in children from 6- to 8-year old is not demonstrated by this study. However, descriptive statistics show a small difference in 7- and 8-year old's justifications for speaker's non-literal statement, namely; 8-year olds gave more justifications which require a higher-order Theory of Mind than 7-year olds. Moreover 10.65% of the justifications of 8-year olds were that the speaker was being ironic. These results indicate that there is a qualitative difference between 7- and 8-year old children in their understanding of irony, suggesting that there is a development in irony understanding in children from 7- to 8-year old. This observation may support the hypothesis that older children perform more adult-like than younger children at irony understanding.

Another interesting finding was that almost no justification for speaker's non-literal statement refers to intonation (1.37%). Various studies have explored children's use of intonation to understand irony, however, the findings are conflicting. Some researchers argue that children make little to no use of intonation to understand irony (Ackerman (1983); Demorest et al. (1984); Winner & Leekam (1991); Sullivan et al. (1995)). On the other hand, Happé (1995) and de Groot et al. (1995) argue that children use intonation to understand irony. Capelli et al. (1990) found that children depend

more heavily on intonation than on context in recognizing sarcasm. Our finding, that children do not point out the intonation of the speaker, suggest that children do not use intonation -or at least that children are not aware of that they use intonation- to understand irony.

The findings of this research seem to indicate that children are not adult-like in irony understanding. Even 8-year olds are not adult-like in their performance. Furthermore, even if a child understands the intention of the speaker, it could be that (s)he is not adult-like in his/her justifications for speaker's non-literal statement. Children do not say that the speaker is being ironic, they give justifications such as: Politeness, Humor, and Opposite. In our study, only 3 out of 35 8-year olds were able to give the justification that the speaker was being ironic. This shows that also children of the age of 8 are not able to understand irony.

Based on our findings, we propose stages of the development in irony understanding (Figure 6.1). The first stage is that children interpret all statements as literal. SEcond, children understand that the meaning could be non-literal. In this stage they are looking for explanations for the non-literal statement, for example lying or making a joke. Finally, the last stage is understanding the speaker's intention for its non-literal statement.

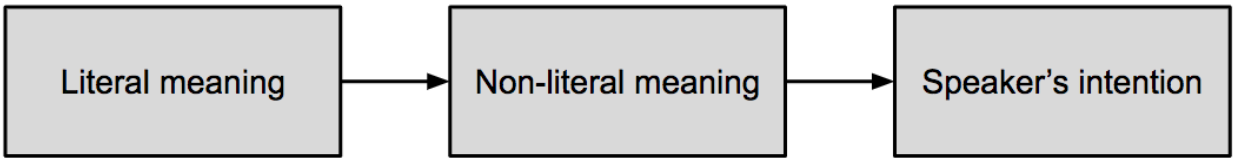

Figure 6.1: Proposed stages of the development of irony understanding.

In conclusion, understanding irony requires cognitive resources. Previous studies in this field used elaborate stories which require more cognitive resources. In our study we used short video fragments with lesser details to minimize the workload, in order for the participants to have more resources available to process an ironic statement. With this method, this experiment could focus on testing irony understanding and its components. However, our results did not prove that by simplifying the task children were able to understand irony at a younger age than previous studies found.

6.2 Limitations and Future Work

This study proved to have some limitations which could be solved in future work. Firstly, it is unfortunate that the study did not include a second-order Theory of Mind task to check if participants who are understand second-order ToM also understand irony. However, due to the length of this irony experiment this would lead to an experiment which would be too long for children. This study could be repeated including a ToM task. To solve the problem that the experiment will take too much time less control items (items with a literal statement) could be shown. In our experiment, 50% of the items were literal statements. Decreasing the amount of control items would not influence the results.

Second, the question about the justification for speaker's non-literal statement was added at the last moment and was not prerecorded. This led to human errors, for example going to the next

question before justification was asked. This resulted in not every participant getting this question for every ironic item. The justification question was only asked in the ironic condition. Additionally, the justification question is a suggestive question; by asking this question, the participant could get the impression that there is something special about the utterance. Since the justification question was only asked in the ironic condition, the risk was there that participants were steered into rethinking their answer. Despite the suggestive nature of this question, a lot of information was received by this question. The results are interesting, but it could be even more impressive when we have more data for this question. Furthermore, we are curious about justification answers of younger and older children, in this study 6-year old children were not asked for their justification for speaker's non-literal statement.

The results of this study indicate that the justifications for the speaker's non-literal statement of children are not similar to adults'. Children say that the ironic speaker is joking, lying or means the opposite. Although this response does not show that participants understand the second-order intention, there is a possibility that children understand irony but do not know how to describe it. If adult participants were asked to explain more without using the words irony or sarcasm, it could be that the children's and adults' responses would be much alike.

Thirdly, in our experiment, the questions about the belief and the emotion of an ironic speaker are formulated negatively or positively. We found that the number of correct answers was higher in items in which the question was formulated positively compared to in items in which the question was negatively formulated. One explanation could be that the belief or emotion in the question does not have a perfect match with the belief or emotion the speaker has according to the participant. In example (6.1) the emotion is a negative emotion, while in example (6.2) the emotion is a positive emotion. For example (6.1) the answer 'NO' was labeled as incorrect and 31% of the answers was incorrectly 'NO'. In example (6.2) the answer 'NO' was labeled as correct and 98% of the answers was correctly 'NO'. We think that this difference is due to the fact that it could be that the mentioned emotion does not represent all the emotions the participant thinks the speaker has. Therefore the participants will lean toward the answer 'NO' unless they are certain it is 'YES'. One solution to avoid this problem is to make the question a multiple choice question like: 'Is the man annoyed or happy?' as did Massaro et al. (2013). In this case you do not have the problem that the answer is not exactly the belief or emotion of the speaker, because the participant has to choose between a positive and a negative evaluative word. Another solution could be a scale like Pexman & Glenwright (2007). They assessed children's perceptions of speaker attitude by a rating scale, a 5-point nice/mean scale.

(6.1) The woman drops her sandwich. The man says: 'Well, you eat neatly!'
Question about emotion: Is the man annoyed that the sandwich was dropped on the table?

(6.2) The man is spilling some coffee. The woman says: 'Damn, you are good at pouring!'
Question about emotion: Is the woman happy that the man spilled some coffee?

An unanticipated finding was that an 8-year old participant answered all the second-order intention of an ironic speaker question incorrectly, but still his/her justification for speaker's non-literal statement was that the speaker was being ironic. This was not what we expected as we predicted that someone who knows that the speaker was ironic would have answered the second-order intention

question also correctly. Although this is just one participant, it is remarkable. Could it be that this person understood the intention of the speaker but did not understand how this question was formulated? The intention question was a long sentence, which could have lead to a misunderstanding of the question, affecting the answers of participants. However in the literal statement condition, exactly the same question was asked and there were no difficulties with understanding the question. This indicates that this question is not too difficult to understand. Still, it could be useful to decrease the number of words of the intention question.

Further research should also be undertaken to investigate the participants' belief and knowledge of the listener. In this experiment the questions were all about the speaker. An example for the belief of the listener is: 'Does the listener think the speaker is saying what (s)he thinks?' 'What does the listener think the speaker means?' This gives insight into whether the participant is able to reason about the thoughts of the listener. In a successful white lie, the listener thinks that the speaker tells the truth. Instead, in a successful ironic statement, the listener thinks that the speaker means the opposite. So, it would also be interesting to study what the participant thinks the belief of the listener is.

The issue of the development of irony understanding is an intriguing one which could be usefully explored in further research. To get more information about how the development proceeds it is interesting to investigate children older than 8-year old. How will irony understanding develop in older children, e.g. 9-, 10-, and 11-year olds? Is it the case that children gradually develop to understand irony? Or is the case that the development of irony understanding proceeds by leaps? The difference in justifications for speaker's non-literal statement from 7-year olds and 8-year olds are small, so when we extend this development our results may suggest that irony understanding will develop gradually and will take some more year before children totally understand irony.

6.3 Larger implications

The findings of this research provide insights in irony understanding in typically developing children. By using short video fragments we minimized the cognitive workload of the participant during the task. By doing so, we make sure that the cognitive sources are not charged with remembering small details like names or locations. We are interested how children and adults who are known to have difficulties with irony and/or ToM, will perform on this task. For example, children and adults with Autism Spectrum Disorder (ASD) have difficulty with Theory of Mind (Baron-Cohen et al., 1985) and understanding irony. It could be possible that by minimizing the cognitive workload for a irony task, children and adults with ASD will understand irony.

Furthermore, in current society, a lot of our communication is not face-to-face communication but by typing and reading written text; such as email, chats and online comments. To develop systems which correctly classify whether a review is positive or negative, we need to take into account that some statements seem to be positive but in reality the message is negative (ironic criticism). Example (6.3) is a review with positive adjectives.

(6.3) 'This restaurant has fantastic food, and the table service is amazing.'

When example (6.3) is a review for a culinary restaurant which has excellent service and fresh food, one could take this review literally. However, when this is a review for a fast food restaurant which has no table service and chewy food, one could take this review ironically. Because obviously the message has not to be taken literal. Nowadays, there is attention for systems which can classify reviews or online comments as ironic or literal. For example, the results of SemEval-2014 (Semantic Evaluation workshop; Nakov et al. (2016)) show that irony is difficult to classify. The task was to classify the sentiment of tweets (positive, neutral, negative), and the performance is measured in $(F1 = (precision * recall)/(precision + recall))$. The best classifier on regular tweets has a higher performance than the best classifier on ironic tweets. This shows that correctly classifying ironic tweets is more difficult than regular tweets. The results of SemEval-2018 Van Hee et al. (2018) show that the performance of the best classifier, which classifies whether a tweet is ironic or not, is almost as high as the performance from the regular tweet classifier from SemEval-2014. However, there is a side note on this performance, since they used a data set of 50% ironic tweets and 50% non-ironic tweets. Samonte et al. (2018) have shown that the results of irony detection is partly based on a balanced data set. They found that in a imbalanced data set (25% ironic and 75% non-ironic tweets) the performance of classifying a tweet as ironic or non-ironic is far less than this performance in a balanced data set (50%, 50%). These studies show that systems are not adult-like, and the question is if that ever will be the case.

In summary, the results of this study are encouraging. Since ironic statements are common in our society, people who are not able to understand irony will have difficulties in social interaction. Therefore it is important to get more insight in irony understanding.

7
Conclusion

Previous studies on irony used elaborate stories and the participant had to remember many details to answer the questions. This requires a lot of cognitive resources. Therefore this makes it difficult to test irony understanding, which also requires resources. In this study, we use short video fragments to minimize the cognitive workload.

The first aim of this study was to examine whether children are adult-like in irony understanding. Even though we minimized the cognitive workload for the experiment, we did not find evidence which led us to believe that by decreasing the cognitive workload for the task, younger children will understand irony. This experiment confirmed that 6- to 8-year old children are not adult-like in irony understanding. In particular children do not understand the second-order intention of a speaker to utter a non-literal statement. For children and adults who have less cognitive resources, for example due to brain injury, this simplified experiment is useful to be able to investigate if they have difficulties with irony understanding.

Furthermore, we found that children gave different justifications for a speaker's non-literal statement than adults. Most of the earlier studies consisted of only close-ended questions. In our study, participants had to answer an open-ended question about the justification for speaker's non-literal statement. The results indicate that adults' justifications are that an ironic speaker is being ironic, but the children's justifications are that the speaker is joking, lying, or means the opposite.

The second aim of this study was to get information about how the development of irony understanding proceeds. From our results it is clear that even 8-year old children are not fully developed in irony understanding. Although children between 6- and 8-years old do not show big differences in their performance, it could be that there is some development present in this age range. In particular, 7-year olds' justifications for speaker's non-literal statement are different from 8-year olds' justification. However, the difference is small and it could be that this development towards adults' justifications takes a long time. The larger question still remains about how and when children fully understand irony.

Ironic statements occur in social interactions, so people who have difficulties with understanding irony, could have difficulties in social interactions. Therefore it is important to get more research done about irony.

References

Ackerman, B. P. (1981). Young children's understanding of a speaker's intentional use of a false utterance. *Developmental Psychology*, *17*(4), 472.

Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology*, *35*(3), 487–508.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, *24*(3), 127–135.

Capelli, C. A., Nakagawa, N., & Madden, C. M. (1990). How children understand sarcasm: The role of context and intonation. *Child Development*, *61*(6), 1824–1841.

Carvalho, P., Sarmento, L., Silva, M. J., & De Oliveira, E. (2009). Clues for detecting irony in user-generated contents: oh...!! it's so easy;-. In *Proceedings of the 1st international cikm workshop on topic-sentiment analysis for mass opinion* (pp. 53–56).

Cornish, E. R., & Wason, P. C. (1970). The recall of affirmative and negative sentences in an incidental learning task. *Quarterly Journal of Experimental Psychology*, *22*(2), 109–114.

de Groot, A., Kaplan, J., Rosenblatt, E., Dews, S., & Winner, E. (1995). Understanding versus discriminating nonliteral utterances: Evidence for a dissociation. *Metaphor and Symbol*, *10*(4), 255–273.

Demorest, A., Meyer, C., Phelps, E., Gardner, H., & Winner, E. (1984). Words speak louder than actions: Understanding deliberately false remarks. *Child Development*, 1527–1534.

Demorest, A., Silberstein, L., Gardner, H., & Winner, E. (1983). Telling it as it isn't: Children's understanding of figurative language. *British Journal of Developmental Psychology*, *1*(2), 121–134.

Faraway, J. J. (2016). *Extending the linear model with r: generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child development*, *79*(1), 126–138.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and symbol*, *15*(1-2), 5–27.

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*(2), 227–248.

Happé, F. G. (1995). Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor and symbol*, *10*(4), 275–295.

Kreuz, R. J., & Caucci, G. M. (2007). Lexical influences on the perception of sarcasm. In *Proceedings of the workshop on computational approaches to figurative language* (pp. 1–4).

Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, *124*(1), 3.

Lüdtke, J., Friedrich, C. K., De Filippis, M., & Kaup, B. (2008). Event-related potential correlates of negation in a sentence–picture verification paradigm. *Journal of cognitive neuroscience*, *20*(8), 1355–1370.

Margolin, S. J., & Abrams, L. (2009). Not may not be too difficult: The effects of negation on older adults' sentence comprehension. *Educational Gerontology*, *35*(4), 308–322.

Massaro, D., Valle, A., & Marchetti, A. (2013). Irony and second-order false belief in children: What changes when mothers rather than siblings speak? *European Journal of Developmental Psychology*, *10*(3), 301–317.

McDonald, S., Flanagan, S., Rollins, J., & Kinch, J. (2003). Tasit: A new clinical tool for assessing social perception after traumatic brain injury. *The Journal of head trauma rehabilitation*, *18*(3), 219–238.

Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., & Stoyanov, V. (2016). Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th international workshop on semantic evaluation (semeval-2016)* (pp. 1–18).

Pell, M. D., Monetta, L., Rothermich, K., Kotz, S. A., Cheang, H. S., & McDonald, S. (2014). Social perception in adults with parkinson's disease. *Neuropsychology*, *28*(6), 905.

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*(3), 437–471.

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

Poletti, M., Enrici, I., Bonuccelli, U., & Adenzato, M. (2011). Theory of mind in parkinson's disease. *Behavioural brain research*, *219*(2), 342–350.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European journal of developmental psychology*, *1*(2), 127–152.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Samonte, M. J. C., Dollete, C. J. T., Capanas, P. M. M., Flores, M. L. C., & Soriano, C. B. (2018). Sentence-level sarcasm detection in english and filipino tweets. In *Proceedings of the 4th international conference on industrial and business engineering* (pp. 181–186).

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, *45*(13), 3054–3067.

Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. *Philosophy*, *3*, 143–184.

Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology*, *13*(2), 191–204.

Van Hee, C., Lefever, E., & Hoste, V. (2018). Semeval-2018 task 3: Irony detection in english tweets. In *Proceedings of the 12th international workshop on semantic evaluation* (pp. 39–50).

Verzani, J. (2014). *Using r for introductory statistics*. Chapman and Hall/CRC.

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

Wimmer, H., Gruber, S., & Perner, J. (1984). Young children's conception of lying: Lexical realism—moral subjectivism. *Journal of Experimental Child Psychology*, *37*(1), 1–30.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Winner, E. (1997). *The point of words: Children's understanding of metaphor and irony*. 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*(2), 257–270.

Woodruff, G., Premack, D., & Kennel, K. (1978). Conservation of liquid and solid quantity by the chimpanzee. *Science*, *202*(4371), 991–994.

Appendices

A Practice and Test Items

The experiment consists of two practice items and 16 test items. For every test situation there are two video fragments; literal statement and ironic statements. *Q2-meaning* has two versions; positive and negative formulated.

Q3-justification was only asked in the ironic statement condition, none of the 6-year old children received this question.

Practice A: Snow

The man and woman are sitting at a table, behind them is a big window with the garden. In the garden there is a lot of snow. The woman says:

Literal statement		Wow, wat ligt er veel sneeuw! *Wow, there is a lot of snow!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	Bedoelt de vrouw dat er weinig sneeuw ligt? *Does the woman mean that there is little snow?*
Q	**Fact**	Ligt er weinig sneeuw? *Is there little snow?*
Q	**Fact**	Is de sneeuw wit. *Is the snow white?*

Practice B: Stairs

The man and woman are standing at the bottom of a long stairs. The man says:

Literal statement		Zo hé, dat is een lange trap! *Oh dear, that is a long staircase!*
Q1	**Intention**	De man zegt tegen de vrouw: `[Literal statement]` Wil hij dat de vrouw dat gelooft? *The man says to the woman:* `[...]` *Does he want the woman to believe this?*
Q2	**Meaning**	Bedoelt de man dat de trap lang is? *Does the man mean that the staircase is long?*
Q	**Fact**	Is de trap lang? *Is the staircase long?*
Q	**Fact**	Heeft de trap twee treden? *Does the staircase have two steps?*

Item 01: Pie

The man and woman are sitting at a table. The man is cutting a pie. He gives the woman a small piece and himself a really big one. The woman says:

Literal statement		Jeetje, wat heb jij de taart oneerlijk verdeeld! *Gee, you divided the pie unequally!*
Ironic statement		Jeetje, wat heb jij de taart eerlijk verdeeld! *Gee, you divided the pie equally!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal or Ironic statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat de man de taart eerlijk heeft verdeeld? *Does the woman mean that the man divided the pie equally?*
	Negative	Bedoelt de vrouw dat de man de taart oneerlijk heeft verdeeld? *Does the woman mean that the man divided the pie unequally?*
(Q3	**Justification)**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw dat ze genoeg taart heeft gekregen? *Does the woman think that she got enough pie?*
Q5	**Emotion**	Is de vrouw blij met hoe de taart is verdeeld? *Is the woman happy about how the pie is divided?*

Item 02: Jumper

The man and woman are standing. The man pulls his jumper on backwards. The woman says:

Literal statement		Nou zeg, jij ziet er belachelijk uit! *Well, you look ridiculous!*
Ironic statement		Nou zeg, jij ziet er fantastisch uit! *Well, you look awesome!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal or Ironic statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat de man er fantastisch uit ziet? *Does the woman mean that the man looks awesome?*
	Negative	Bedoelt de vrouw dat de man er belachelijk uit ziet? *Does the woman mean that the man looks ridiculous?*
(Q3	**Justification)**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw dat de man het vest goed heeft aangetrokken? *Does the woman think that the man has pulled his jumper the right way on?*
Q5	**Emotion**	Is de vrouw ontevreden over hoe de man eruit ziet? *Is the woman unsatisfied how the man looks?*

Item 03: Coffee

The man and woman are sitting at a table. The man is pouring coffee in a cup, however he is spilling some coffee.

Literal statement		Zo hé, wat kun jij slecht inschenken! *Oh dear, you are bad at pouring!*
Ironic statement		Zo hé, wat kun jij goed inschenken! *Oh dear, you are good at pouring!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal or Ironic statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat de man goed kan inschenken? *Does the woman mean that the man is good at pouring?*
	Negative	Bedoelt de vrouw dat de man slecht kan inschenken? *Does the woman mean that the man is bad at pouring?*
(Q3	**Justification)**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw het onhandig dat de man koffie heeft gemorst? *Does the woman think it is clumsy that the man spilled some coffee?*
Q5	**Emotion**	Is de vrouw blij dat de man koffie heeft gemorst? *Is the woman happy that the man spilled some coffee?*

Item 04: Sandwich

The man and woman are sitting at a table. The woman drops her sandwich. The man says:

Literal statement		Wat eet jij slordig, zeg! *Well, you eat sloppy!*
Ironic statement		Wat eet jij netjes, zeg! *Well, you eat neatly!*
Q1	**Intention**	De man zegt tegen de vrouw: `[Literal or Ironic statement]` Wil hij dat de vrouw dat gelooft? *The man says to the woman:* `[...]` *Does he want the woman to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de man dat de vrouw netjes eet? *Does the man mean that the woman eats neatly!*
	Negative	Bedoelt de man dat de vrouw slordig eet? *Does the man mean that the woman eats sloppy?*
(Q3	**Justification)**	Waarom zegt hij dit? *Why does he say this?*
Q4	**Belief**	Vindt de man het slecht dat de vrouw het brood op tafel heeft laten vallen? *Does the man think it is bad that the woman drops her sandwich on the table?*
Q5	**Emotion**	Is de man chagrijning dat het brood op tafel is gevallen? *Is the man annoyed that the sandwich was dropped on the table?*

Item 05: Handstand

The man and woman are standing. The woman is making a handstand, however that handstand is not a proper one. The man says:

Literal statement		Dat was een verschrikkelijke handstand! *That was a horrible handstand!*
Ironic statement		Dat was een geweldige handstand! *That was a fantastic handstand!*
Q1	**Intention**	De man zegt tegen de vrouw: `[Literal or Ironic statement]` Wil hij dat de vrouw dat gelooft? *The man says to the woman:* `[...]` *Does he want the woman to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de man dat het een geweldige handstand was? *Does the man mean that the handstand was fantastic.*
	Negative	Bedoelt de man dat het een verschrikkelijke handstand was? *Does the man mean that the handstand was horrible?*
(Q3	**Justification)**	Waarom zegt hij dit? *Why does he say this?*
Q4	**Belief**	Vindt de man dat de vrouw een goede handstand heeft gemaakt? *Does the man think that the woman did a good handstand?*
Q5	**Emotion**	Is de man tevreden over hoe de handstand is gemaakt? *Is the man satisfied how the handstand was done?*

Item 06: Tower blocks

The man and woman are sitting at a table and they are building a tower out of wooden blocks. The woman makes the tower to fall over. The man says:

Literal statement		Jij bent slecht in dit spel, zeg! *Well, you are bad at this game!*
Ironic statement		Jij bent goed in dit spel, zeg! *Well, you are good at this game!*
Q1	**Intention**	De man zegt tegen de vrouw: `[Literal or Ironic statement]` Wil hij dat de vrouw dat gelooft? *The man says to the woman:* `[...]` *Does he want the woman to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de man dat de vrouw goed in het spelletje is? *Does the man mean that the woman is good in the game?*
	Negative	Bedoelt de man dat de vrouw slecht in het spelletje is? *Does the man mean that the woman is bad in the game?*
(Q3	**Justification)**	Waarom zegt hij dit? *Why does he say this?*
Q4	**Belief**	Vindt de man het slecht dat de vrouw de blokjes heeft laten vallen? *Does the man think it is bad that the woman dropped the blocks?*
Q5	**Emotion**	Is de man ontevreden met hoe het spelletje gaat? *Is the man unsatisfied about how the game is going?*

Item 07: Tennis

The man and woman are playing tennis. The woman drops the ball. The man says:

Literal statement		Nou zeg, jij bent hier slecht in! *Oh dear, you are bad at this!*
Ironic statement		Nou zeg, jij bent hier goed in! *Oh dear, you are good at this!*
Q1	**Intention**	De man zegt tegen de vrouw: `[Literal or Ironic statement]` Wil hij dat de vrouw dat gelooft? *The man says to the woman:* `[...]` *Does he want the woman to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de man dat de vrouw goed kan tennissen? *Does the man means that the woman is good at tennis?*
	Negative	Bedoelt de man dat de vrouw slecht kan tennissen? *Does the man means that the woman is bad at tennis?*
(Q3	**Justification)**	Waarom zegt hij dit? *Why does he say this?*
Q4	**Belief**	Vindt de man dat de vrouw goed heeft getennist? *Does the man think that the woman played tennis well?*
Q5	**Emotion**	Is de man ontevreden over hoe het tennissen gaat? *Is the man unsatdisfied about the tennis game?*

Item 08: Cup

The man and woman are sitting at a table. The woman knocks a cup from the table. The man says:

Literal statement		Dat was erg onhandig van je! *That was really clumsy of you!*
Ironic statement		Dat was erg handig van je! *That was really handy of you!*
Q1	**Intention**	De man zegt tegen de vrouw: `[Literal or Ironic statement]` Wil hij dat de vrouw dat gelooft? *The man says to the woman:* `[...]` *Does he want the woman to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de man dat het handig is van de vrouw? *Does the man means that it is handy of the woman?*
	Negative	Bedoelt de man dat het onhandig is van de vrouw? *Does the man means that it is unhandy of the woman?*
(Q3	**Justification)**	Waarom zegt hij dit? *Why does he say this?*
Q4	**Belief**	Vindt de man het onhandig dat de vrouw het glas heeft omgegooid? *Does the man think it is clumsy that the woman knocks over a cup?*
Q5	**Emotion**	Is de man blij dat het glas is omgevallen? *Is the man happy that the cup is knocked over?*

Item 09: Computer

The man and woman are sitting at a table and the woman is working on the computer. She says:

Literal statement		Vervelend, de computer is kapot! *Awful, the computer is broken!*
Ironic statement		Geweldig, de computer is kapot! *Great, the computer is broken!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal or Ironic statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat het geweldig is dat de computer kapot is? *Does the woman mean that it is great that the computer is broken?*
	Negative	Bedoelt de vrouw dat het vervelend is dat de computer kapot is? *Does the woman mean that it is awful that the computer is broken?*
(Q3	**Justification)**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw dat het een goede computer is? *Does the woman think that the computer is good?*
Q5	**Emotion**	Is de vrouw chagrijnig dat de computer stuk is? *Is the woman annoyed that the computer is broken?*

Item 10: Pasta

The man and woman are sitting at a table. The woman is pulling a bag of pasta out of her handbag. The pasta bag is broken and there is pasta falling out of it. The woman says:

Literal statement		Jammer dat de zak stuk is! *It is a pity that the package is broken!*
Ironic statement		Fijn dat de zak stuk is! *It is great that the package is broken!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal or Ironic statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat het fijn is dat de zak stuk is? *Does the woman mean that it is nice that the package is broken?*
	Negative	Bedoelt de vrouw dat het jammer is dat de zak stuk is? *Does the woman mean that it is a pity that the package is broken?*
(Q3	**Justification)**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw dat het een goede zak is? *Does the woman think the package is great?*
Q5	**Emotion**	Is de vrouw chagrijnig dat de zak stuk is? *Is the woman annoyed about the broken package?*

Item 11: Swing

The man and woman are standing near by a broken swing. The woman says:

Literal statement		Jammer, de schommel is stuk! *It's a pity, the swing is broken!*
Ironic statement		Geweldig, de schommel is stuk! *Awesome, the swing is broken!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal or Ironic statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat het geweldig is dat de schommel stuk is? *Does the woman mean that it is awesome that the swing is broken?*
	Negative	Bedoelt de vrouw dat het jammer is dat de schommel stuk is? *Does the woman mean that it is a pity that the swing is broken?*
(Q3	**Justification)**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw dat het een slecht schommel is? *Does the woman think that it is a bad swing?*
Q5	**Emotion**	Is de vrouw blij dat de schommel stuk is? *Is the woman happy that the swing is broken?*

Item 12: Toy

The man and woman are playing with a toy car. The wheel is broken. The woman says:

Literal statement		Naar zeg, het autootje is stuk! *How unfortunate, the toy car is broken!*
Ironic statement		Fijn zeg, het autootje is stuk! *How nice, the toy car is broken!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal or Ironic statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat het fijn is dat het autootje stuk is? *Does the woman mean that it is nice that the toy car is broken?*
	Negative	Bedoelt de vrouw dat het naar is dat het autootje stuk is? *Does the woman mean that it is unfortunate that the toy car is broken?*
(Q3	**Justification)**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw dat het een slecht autootje is? *Does the woman think the toy car is no good?*
Q5	**Emotion**	Is de vrouw blij dat het autootje stuk is? *Is the woman happy that the toy car is broken?*

Item 13: Puzzle

The man and woman are making a jigsaw puzzle. One piece is missing. The woman says:

Literal statement		Wat vervelend dat er een stukje mist! *How sad that there is a piece missing!*
Ironic statement		Wat leuk dat er een stukje mist! *How great that there is a piece missing!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal or Ironic statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat het leuk is dat er een stukje mist? *Does the woman mean that it is great that there is a missing piece?*
	Negative	Bedoelt de vrouw dat het vervelend is dat er een stukje mist? *Does the woman mean that it is sad that there is a missing piece?*
(Q3	**Justification)**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw dat het een slechte puzzel is? *Does the woman think it is a bad puzzle?*
Q5	**Emotion**	Is de vrouw ontevreden met dat er een stukje mist? *Is the woman unsatisfied that there is a missing piece?*

Item 14: Trash bag

The man is lift up a trash bag, which is broken. All the trash falls out of the bag. The woman says:

Literal statement		Wat super onhandig dat de zak scheurt! *That's just terrible, the bag ripped!*
Ironic statement		Wat super handig dat de zak scheurt! *That's just great, the bag ripped!*
Q1	**Intention**	De vrouw zegt tegen de man: `[Literal or Ironic statement]` Wil ze dat de man dat gelooft? *The woman says to the man:* `[...]` *Does she want the man to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de vrouw dat het handig is dat de zak scheurt? *Does the woman mean that it is great that the bag ripped?*
	Negative	Bedoelt de vrouw dat het onhandig is dat de zak scheurt? *Does the woman mean that it is terrible that the bag ripped?*
(Q3	**Justification)**	Waarom zegt ze dit? *Why does she say this?*
Q4	**Belief**	Vindt de vrouw dat het een slechte vuilniszak is? *Does the woman think it is a bad trash bag?*
Q5	**Emotion**	Is de vrouw ontevreden met dat de zak scheurt? *Is the woman unsatisfied that the bag ripped?*

Item 15: Desk

The man and woman are sitting at a messy desk. The man says:

Literal statement		Wooh, dit is rommelig! *Woow, this is a mess!*
Ironic statement		Wooh, dit is opgeruimd! *Woow, this is tidy!*
Q1	**Intention**	De man zegt tegen de vrouw: `[Literal or Ironic statement]` Wil hij dat de vrouw dat gelooft? *The man says to the woman:* `[...]` *Does he want the woman to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de man dat het opgeruimd is? *Does the man mean that it is tidy?*
	Negative	Bedoelt de man dat het rommelig is? *Does the man mean that it is messy?*
(Q3	**Justification)**	Waarom zegt hij dit? *Why does he say this?*
Q4	**Belief**	Vindt de man dat het bureau netjes is? *Does the man think the desk is tidy?*
Q5	**Emotion**	Is de man blij met de rommel? *Is the man happy with the mess?*

Item 16: Chair

The man is going to sit down on a chair. The chair is broken. The man says:

Literal statement		Dit is een slechte stoel zeg! *Well, this is a useless chair!*
Ironic statement		Dit is een goede stoel zeg! *Well, this is a nice chair!*
Q1	**Intention**	De man zegt tegen de vrouw: `[Literal or Ironic statement]` Wil hij dat de vrouw dat gelooft? *The man says to the woman:* `[...]` *Does he want the woman to believe this?*
Q2	**Meaning**	
	Positive	Bedoelt de man dat het een goede stoel is? *Does the man mean that it is a good chair?*
	Negative	Bedoelt de man dat het een slechte stoel is? *Does the man mean that it is a bad chair?*
(Q3	**Justification)**	Waarom zegt hij dit? *Why does he say this?*
Q4	**Belief**	Vindt de man dat hij op een goede stoel is gaan zitten? *Does the man think he is sitting on a bad chair?*
Q5	**Emotion**	Is de man blij met de stoel? *Is the man happy with the chair?*

B Video Fragments

Information about statement in the video fragments. The number of words include interjections. From each item consist two fragments; one with the ironic statement (I) and one with the literal statement (L).

	Scenario	**Victim of irony**	**Gender of Speaker**	**Number of Words**	**Length of Fragment (seconds)**
Practice A	Snow	-	Female	6	07
Practice B	Stairs	-	Male	7	04
01	Pie	Listener	Female	8	I: 21 L: 20
02	Jumper	Listener	Female	7	I: 11 L: 11
03	Coffee	Listener	Female	7	I: 11 L: 13
04	Sandwich	Listener	Male	5	I: 06 L: 05
05	Handstand	Listener	Male	5	I: 06 L: 07
06	Tower Blocks	Listener	Male	7	I: 08 L: 08
07	Tennis	Listener	Male	7	I: 06 L: 06
08	Cup	Listener	Male	6	I: 11 L: 08
09	Computer	Situation	Female	5	I: 06 L: 07
10	Pasta	Situation	Female	6	I: 06 L: 06
11	Swing	Situation	Female	5	I: 06 L: 06
12	Toy	Situation	Female	6	I: 08 L: 08
13	Puzzle	Situation	Female	7	I: 06 L: 07
14	Trash bag	Situation	Female	7	I: 07 L: 07
15	Desk	Situation	Male	4	I: 05 L: 06
16	Chair	Situation	Male	6	I: 08 L: 09

C Experiment Lists

There were four different lists of the experiment, see below. Ironic and literal means that the criticism was ironic or literal. 'Q2_pos' and 'Q2_neg' means that comprehension question *Q2-meaning* was positively or negatively formulated.

Nr	Item	List 1	List 2	List 3	List 4
04	Sandwich	ironic - Q2_pos	ironic - Q2_neg	literal - Q2_pos	literal- Q2_neg
05	Handstand	ironic - Q2_pos	ironic - Q2_neg	literal - Q2_pos	literal- Q2_neg
06	Tower blocks	ironic - Q2_neg	literal - Q2_pos	literal- Q2_neg	ironic - Q2_pos
01	Pie	ironic - Q2_neg	literal - Q2_pos	literal- Q2_neg	ironic - Q2_pos
02	Jumper	literal - Q2_pos	literal- Q2_neg	ironic - Q2_pos	ironic - Q2_neg
07	Tennis	literal - Q2_pos	literal- Q2_neg	ironic - Q2_pos	ironic - Q2_neg
08	Cup	literal- Q2_neg	ironic - Q2_pos	ironic - Q2_neg	literal - Q2_pos
03	Coffee	literal- Q2_neg	ironic - Q2_pos	ironic - Q2_neg	literal - Q2_pos
15	Desk	ironic - Q2_pos	ironic - Q2_neg	literal - Q2_pos	literal- Q2_neg
09	Computer	ironic - Q2_pos	ironic - Q2_neg	literal - Q2_pos	literal- Q2_neg
10	Pasta	ironic - Q2_neg	literal - Q2_pos	literal- Q2_neg	ironic - Q2_pos
11	Swing	ironic - Q2_neg	literal - Q2_pos	literal- Q2_neg	ironic - Q2_pos
12	Toy	literal - Q2_pos	literal- Q2_neg	ironic - Q2_pos	ironic - Q2_neg
16	Chair	literal - Q2_pos	literal- Q2_neg	ironic - Q2_pos	ironic - Q2_neg
13	Puzzle	literal- Q2_neg	ironic - Q2_pos	ironic - Q2_neg	literal - Q2_pos
14	Trash bag	literal- Q2_neg	ironic - Q2_pos	ironic - Q2_neg	literal - Q2_pos

D Categories of Justifications

The table in this appendix shows all the answers of the participants per age group at question *Q3-justification* with their classified category. This question was not asked to 6-year old children.

Sometimes the participant said *he* or *she*, we rephrased this as *listener (luisteraar)* and *speaker (spreker)* to be more clear.

Category	Waarom zegt hij/zij dat?	7-y.o.	8-y.o.	Adults
Ironic	ik weet nog iets anders dan sarcastisch: humor			1
Ironic	ironisch			1
Ironic	laat ik nu iets anders zeggen dan sarcastisch. Het is raar om te zeggen dat het een bende is			1
Ironic	sarcastisch		6	97
Ironic	sarcastisch met irritatie			1
Ironic	sarcastisch, als grapje			1
Ironic	sarcastisch, als je het meent heb je een hogere stem		1	
Ironic	sarcastisch, je gebruikt het tegenovergestelde om je uitspraak kracht bij te zetten			1
Ironic	sarcastisch, want hij is geirriteerd			1
Ironic	sarcastisch, ze zegt iets wat ze juist niet vindt			1
Ironic	sarcastisch, zonder irritatie			1
Ironic	soms zeg je iets wat je niet bedoelt. dat is dan sarcastisch			1
Ironic	spottend			2
Ironic	spreker baalt en ze zegt het sarcastisch			1
Ironic	spreker bedoelt het sarcastisch		16	
Ironic	spreker bedoelt juist niet. sarcastisch			1
Ironic	spreker impliceert dat er iets aan gedaan moet worden. spottend			1
Ironic	spreker zegt het tegenovergestelde. hij is sarcastisch			1
White Lie	anders is het zielig voor de luisteraar	4		
White Lie	anders maakt spreker luisteraar boos	1		
White Lie	anders wordt de luisteraar heel boos	1		
White Lie	dan denkt luisteraar dat het minder erg is		1	
White Lie	dan denkt luisteraar dat het niet zo er is		1	
White Lie	grapje. spreker wil de luisteraar niet boos maken	1		
White Lie	‘maar niet heus’, spreker wil de luisteraar niet beledigen maar wel een beetje	1		
White Lie	om het zo te zeggen dat is minder vervelend voor de luisteraar		1	
White Lie	om luisteraar te steunen	1		
White Lie	spreker bedoelt juist niet. De spreker wil niet dat de luisteraar boos wordt		1	
White Lie	spreker liegt	8		

Category	Waarom zegt hij/zij dat?	7-y.o.	8-y.o.	Adults
White Lie	spreker meent het niet, maar anders moet luisteraar huilen	1		
White Lie	spreker probeert de luisteraar niet boos/verdrietig te maken		1	
White Lie	spreker probeert luisteraar niet verdrietig te maken		1	
White Lie	spreker wil luisteraar niet beledigen en wil aardig doen	1		
White Lie	spreker wil luisteraar niet boos/verdrietig maken		1	
White Lie	spreker wil luisteraar niet kwetsen		4	
Politeness	anders zou spreker het heel boos zeggen	1		
Politeness	beleefder		1	
Politeness	beleefder dan te schreeuwen		1	
Politeness	betere manier van zeggen		2	
Politeness	omdat spreker niet boos wil worden		1	
Politeness	‘slecht’ is een onaardig woord	1		
Politeness	speciaal stemmetje, dit is aardiger dan ‘lelijk’		1	
Politeness	spreker bedoelt andersom, maar ‘slecht’ komt lullig over	1		
Politeness	spreker bedoelt juist niet, beleefder		1	
Politeness	spreker bedoelt tegenovergestelde. Mensen doen dat vaak, omdat dat beleefder is dan: pfff slechte stoel		1	
Politeness	spreker is bozig, maar het is niet netjes om een schelddwoord te gebruiken		1	
Politeness	spreker mag niet schelden		1	
Politeness	spreker wil aardig zijn tegen de luisteraar		1	
Politeness	spreker zegt dit om aardig te zijn	4		
Opposite	‘geweldig’ betekent soms ‘jammer’	2		
Opposite	aandacht vestigen op het slechte - omgekeerde psychologie			1
Opposite	als je het boos zegt dan bedoel je het andersom	1		
Opposite	als ze boos zijn zeggen sommige mensen het tegenovergestelde		1	
Opposite	ander woord voor ‘nee he’		1	
Opposite	andere manier om te zeggen: ‘nee he’		1	
Opposite	bedoelt andersom	1		
Opposite	de spreker wil een punt maken en zegt het daarom omgekeerd			1
Opposite	sommige mensen zeggen leuk als ze stom bedoelen		1	

Category	Waarom zegt hij/zij dat?	7-y.o.	8-y.o.	Adults
Opposite	soms betekent dat juist niet	3		
Opposite	soms kun je omgekeerd bedoelen	1		
Opposite	soms omgekeerd	1		
Opposite	soms omgekeerde, bijvoorbeeld als het regent zeggen sommige mensen: 'lekker weertje'		1	
Opposite	soms zeggen mensen het omgekeerde		1	
Opposite	soms zijn mensen boos en dan zeggen ze het omgekeerde		1	
Opposite	spreker bedoelt andersom	1	5	
Opposite	spreker bedoelt andersom. ze kijkt er raar bij		1	
Opposite	spreker bedoelt juist 'stom'		1	
Opposite	spreker bedoelt juist niet	25	53	
Opposite	spreker bedoelt juist niet, de stem klinkt dan raar		1	
Opposite	spreker bedoelt juist niet, dit is een betere manier van zeggen		2	
Opposite	spreker bedoelt juist niet. Mama zegt ook wel eens: 'geweldig dat je nog een keer het glas laat vallen'		1	
Opposite	spreker bedoelt juist niet. Mijn moeder zegt dat ook wel eens		1	
Opposite	spreker bedoelt juist niet. Mijn vader zegt dat ook wel eens zo	1		
Opposite	spreker bedoelt juist niet, spreker kijkt er raar bij		2	
Opposite	spreker bedoelt juist niet, ze zucht namelijk		1	
Opposite	spreker bedoelt maar niet heus	6		
Opposite	spreker bedoelt niet-fijn		1	
Opposite	spreker bedoelt omgekeerde	5	5	
Opposite	spreker bedoelt tegenovergestelde	1	2	
Opposite	spreker heeft een speciaal stemmetje, spreker bedoelt juist niet		1	
Opposite	spreker is boos en dan zeg je soms andersom		1	
Opposite	spreker is bozig. dus zegt hij 'fijn'		1	
Opposite	spreker is chagrijnig en daarom zegt spreker tegenovergestelde	1		
Opposite	spreker meent het niet	6	3	
Opposite	spreker wil benadrukken dat het een slechte computer is		1	
Opposite	spreker zegt: 'geweeeellldig' dan bedoelt ze juist niet	1		

Category	Waarom zegt hij/zij dat?	7-y.o.	8-y.o.	Adults
Opposite	spreker zegt: ‘zooo’ dat betekent juist niet	1		
Opposite	‘zoooo hee wat gooeed’. Spreker bedoelt juist niet	1		
Humor	anders is het te serieus		1	
Humor	boos grapje	1	2	
Humor	geintje	5		
Humor	grapje	18	10	4
Humor	grapje wat grote mensen vaak maken		1	
Humor	grapje. mama doet dat ook weleens		1	
Humor	omgekeerd grapje	1		
Humor	soort grapje	5		
Humor	spreker verbloemt het onaardige door een grapje			1
Humor	spreker zegt ‘handig’ zodat het niet heel serieus is. en anders schaamt de luisteraar zich		1	
Humor	stom grapje	1		
Proverb	manier van zeggen		4	
Proverb	spreekwoord		9	
Proverb	uitdrukking	4	8	
Intonation	klinkt bozer & dan heeft luisteraar het sneller door		1	
Intonation	speciaal stemmetje		3	
Intonation	ze zegt het zo van: zooo heeee		1	
Other	dan begrijpt luisteraar het makkelijker		1	
Other	de spreker hoopt dat de luisteraar wel door heeft dat hij er belachelijk uitziet			1
Other	extra nadruk op leggen			1
Other	het leek alsof de spreker het niet goed vond	1		
Other	ik zie dat aan het gezicht van de spreker		1	
Other	ik zou ook niet boos zijn. het ging per ongeluk		1	
Other	luisteraar morst		1	
Other	om luisteraar te sarren			1
Other	spreker bedoelt dat ze boos is		1	
Other	spreker is boos		3	
Other	spreker trekt er een stom hoofd bij		1	
Other	spreker vindt het jammer dat het spelletje niet goed gaat			1
Other	spreker wil luisteraar belachelijk maken		2	
Other	spreker zegt het zo. om zijn irritatie te uiten			3
Literal	spreker zei dat luisteraar er mooi uitzag. en spreker meende het	1		
No Answer	weet ik niet	29	28	