Cartoons and comprehension: The effect of visual context on children's sentence processing

by

Rebecca Cooper

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

SPEECH-LANGUAGE PATHOLOGY

Department of Communication Sciences & Disorders
University of Alberta

Abstract

Spoken language comprehension is a complex task that involves taking into account many types of information at the same rate as speech unfolds. It is established that adults use both non-linguistic cues, such as visual context, as well as linguistic cues, such as semantic animacy, to maximize successful interpretation of the input during language processing. Previous research has shown that adults can override their usual assumptions of animacy and interpret inanimate entities as the actors in a sentence when provided with a fictional context. To date no research has looked at whether this flexibility of processing extends to children.

This study examined how semantic animacy and visual features of the antecedent nouns constrain children's and adults' interpretations of ambiguous pronouns during spoken language comprehension. The semantic animacy of the subject and object nouns varied (i.e., animate animals vs. inanimate objects). We manipulated visual animacy by adding eyes and mouths to the inanimate images in half of the trials (e.g., for "There is the couch that the bunny phoned" the image of the couch included facial features). By tracking participants' eye movements to the images of the subject and object of the preceding sentence we observed how they interpreted the ambiguous agentive pronoun, "he" (e.g., "He was excited to go to the movie that night"). We also collected offline data in the form of verbal responses to pronoun antecedent questions.

Eye-tracking data revealed that both children and adults used semantic animacy during pronoun resolution. However time-course and sentence condition differences between participant groups suggest subtle differences between adults and children online. Verbal responses showed that both groups preferred animate nouns as antecedents in final interpretations. Children also displayed main effects of visual context in eye-tracking data, when adults only displayed interactions. Yet pronoun resolution was affected by visual context in both groups' offline verbal responses. Inanimate nouns were more likely to be considered antecedents when corresponding images were visually animate (i.e., contained facial features). Findings from this study illustrate children's ability to use visual context during pronoun resolution, as well as illustrate similarities and differences in processing in comparison to adults.

Preface

This thesis is an original work by Rebecca Cooper. The research project, of which this thesis project is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name “Manipulating the Constraint of Animacy with Visual Context”, No. Pro00049753, September 18, 2015.

Acknowledgements

I would first like to thank my thesis supervisors Dr. Monique Charest of the Department of Communication Sciences and Disorders, and Dr. Juhani Järvikivi of the Department of Linguistics at the University of Alberta. Both showed tireless patience with me over the span of this research project. Dr. Charest has been an invaluable mentor to me on many subjects, both academically and clinically in my journey to becoming a speech-language pathologist. Dr. Järvikivi has been instrumental in shaping how I have considered language and the human mind over the past five years. Both have pushed me to reach for the highest standards of excellence in everything I do.

I would also like to thank Dr. Carol Boliek of the Department of Communication Sciences and Disorders at the University of Alberta for serving as the third committee member, and Dr. Elena Nicoladis of the Department of Psychology at the University of Alberta for serving as the external reviewer for this thesis. I am gratefully indebted to them for their very valuable comments on this thesis.

Thank you to the preschools and daycares of Edmonton as well as the individual families that participated in the study. Research such as this is not possible without willing participants. Thank you also to volunteers Megan Wilson, Jillian Smith, Katrina Bartel, Thomas Pasterfield, and Paulo Arago for their help with data collection.

Thank you to Social Sciences and Humanities Research Council of Canada and the University of Alberta for the Joseph-Armand Bombardier Canada Graduate Scholarship and the Walter H. Johns Graduate Fellowship respectively.

Finally, I must express my very profound gratitude to my parents and siblings, my good friends Kelsey McCarroll, Genevieve Roberts, and others, and especially to my partner, Sean Cunningham, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Table of Contents

List of Tables

List of Figures

Introduction

The ability to participate in a conversation or follow spoken directions requires immediate online comprehension, as spoken language is a transient signal that disappears as quickly as it is presented. A listener must not only understand the individual words presented, but also how they relate to each other, how they relate to the context they are presented in, and what the intent behind the message is. All of this information is dealt with at the same rate as speech unfolds, so that the listener can react in a timely manner. Comprehension research in psycholinguistics aims to understand how humans can succeed at such a complex task so quickly and successfully.

Constraint-based theories represent one prominent approach to language comprehension. As a class, these theories share the common position that all possible linguistic and non-linguistic cues are used simultaneously during comprehension to facilitate successful interpretation of the input (e.g., Coco & Keller, 2015; Ferreira, Foucart & Engelhardt, 2013; Spivey-Knowlton & Sedivy, 1995; Trueswell, Tanenhaus & Kello, 1993). Linguistic cues include phonetic, syntactic, and semantic information embedded in the speech signal. Non-linguistic cues include information such as speaker intent and the visual context the listener sees (Coco & Keller, 2015; Nappa, Wessel, McEldoon, Gleitman & Trueswell, 2009). Constraint-based models of comprehension have been investigated and corroborated extensively with adult listeners. However, it is currently unknown how and when children learn to use linguistic and visual cues in an adult way. It is unclear when children begin to integrate multiple types of information to process the incoming language, when the processing becomes adult-like, and what types of cognitive processes are behind such abilities. Such knowledge will eventually lead to a stronger understanding of typical language development, and will provide a foundation for investigations into atypical language development.

In this study, we examined the influence of semantic animacy – a linguistic cue indicating whether an entity is alive and volitional in the real world– on young children's language comprehension, and how this may be modulated by the non-linguistic cue of visual animacy – that is, whether the referent has animate features, such as a face. This study used the Visual World Paradigm, an experimental method that tracks eye gaze to images during spoken language comprehension to gain insight into language processing in real time (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). Sentences containing inanimate objects taking on animate roles (e.g., a TV acting as an agent) were presented auditorily while cartoon images of the objects were shown with facial features (e.g., an image of a TV with eyes and a mouth). Participants' eye movements were tracked to measure how the experimental manipulations affected their comprehension processes. Online, eye-tracking data and offline, language interpretation data provided information on how the two types of animacy (semantic and visual) influenced pronoun resolution of an ambiguous pronoun in children and adults. Effects of both semantic animacy and visual animacy on referential processing were found for both children and adults, though with some differences between groups. Overall, there were preferences for semantically and visually animate nouns as pronoun antecedents. However, in the offline language interpretation data, animacy effects appeared in a step-wise fashion. Semantically inanimate nouns were least likely to be considered as antecedents. When given faces, they were more likely to be considered as antecedents, yet animate nouns continued to be most likely considered as antecedents.

This study contributes to our understanding of whether children can immediately integrate visual and linguistic information to guide language interpretation in a similar way to adults, a question that is currently subject to debate (Trueswell, Sekerina, Hill & Logrip, 1999; Snedeker & Trueswell, 2004; Weighall & Altmann, 2011; Zhang & Knoeferele, 2012). Additionally, it investigates the interaction of semantic cues and non-linguistic cues in both children's and adults' language processing. It contributes both to the development of constraint-based models for adult language comprehension and to understanding language acquisition within constraint-based frameworks.

Background

Relative Clauses

This study used object relative clauses, such as "There is the snake that the bunny phoned at school", as a means to examine the research questions at hand. This is a syntactic form that varies in processing difficulty as a function of the semantic animacy of the noun referents, and is also processed successfully by young children (Clifton et al., 2003; Diessel & Tomasello, 2005; Kidd, Brandt, Lieven and Tomasello, 2007; Lowder & Gordon, 2014; Mak, Vonk & Schriefers, 2002; Traxler, Morris & Seely, 2002; Trueswell, Tanenhaus & Garnsey, 1994).

Relative clauses modify head nouns. In the examples below, the head nouns are indicated in italics. The role of the head noun in the main clause is classified by "embeddedness". Centrally embedded relative clauses modify the subject of the matrix verb, such as in (1) and (2). Right-branching relative clauses modify the object of the matrix verb, such as in (3) and (4). Additionally, the head noun can take the role of the subject or object of the subordinate clause. (1) and (3) are examples of subject relative clauses, while (2) and (4) are examples of object relative clauses. The role of the head noun in the relative clause is indicated by the position of the gap in the relative clause in the examples below.

(1) *The woman* (that ______ was in charge) paid the electrician.

(2) *The woman* (that the boss had delegated ______) paid the electrician.

(3) The woman paid *the electrician* (that ______ was in charge).

(4) The woman paid *the electrician* (that the boss had delegated ______).

The roles of the head noun in the main clause (centrally embedded vs., right-branching) and in the relative clause (subject vs. object) are important determinants of processing demands for adults and children. Originally, object relative clauses were thought to be more difficult to process than subject relative clauses (Ford, 1983; King & Just, 1991), as indicated by longer reading times for object relative clauses. However, when semantic animacy of the head noun is controlled for, ease of processing is equivalent for both types of relative clauses (Clifton et al., 2003; Lowder & Gordon, 2014; Mak et al., 2002; Traxler et al., 2002; Trueswell et al., 1994). For example, Trueswell et al. (1994) found in an eye-tracking experiment that participants had slower reading times with reduced object relative clauses when the head noun was animate in comparison to when the head noun was inanimate (for example, (5)a. vs. (5)b. from Trueswell et al. (1994, p.286)), suggesting greater processing difficulty.

(5) a. The defendant examined by the lawyer turned out to be unreliable.

b. The evidence examined by the lawyer turned out to be unreliable.

The authors suggested that the pattern of sentence-initial animate nouns acting as the thematic agents of sentences is a very strong constraint for listeners and readers. That is, sentences such as (5)a. incorrectly constrain the language comprehension system into predicting a simple active sentence structure as soon as "the defendant" is recognized. When the reader gets to "by the lawyer", they find they have parsed the sentence in error and must reinterpret it. However, for sentences such as (5)b. the thematic (i.e., semantic) constraint necessitating that sentences have animate agents is violated immediately because "evidence" cannot be the agent of the sentence. This violation constrains the possible syntactic frames to only those that allow the thematic roles of themes and patients sentence initially, such as object relative clauses.

This difference in processing patterns based on animacy has also been found with unreduced relative clauses (Clifton, et al., 2003; Lowder & Gordon, 2014). These are clauses containing relative pronouns such as *that*, *who,* etc. (e.g., examples (1) through (4) above). Additionally, Traxler et al. (2002) found that object relative clauses with inanimate head nouns (e.g., "The evidence that the lawyer examined turned out to be unreliable") were processed as easily as subject relative clauses. Mak et al. (2002) replicated these findings in Dutch, and argued that animacy effects may be based in the distribution of relative clause usage. In looking at a corpus of German and Dutch, the authors found that object relative clauses are more likely to appear with an inanimate head noun (e.g., "Because of the investigation, the computer, that the burglars stole, had to remain at the police station for some time"), while subject relative clauses are more likely to appear with an animate head noun (e.g., "Because of the investigation, the burglars, who robbed the occupant, had to stay at the police station for some time")

The history of understanding relative clause processing in children has mirrored that of adults. Initial studies suggested that processing of object relative causes was of particular difficulty for children in comparison to subject relative clauses (Arnon, 2005; Corrêa, 1995; Kidd & Bavin, 2002). However, by testing children on the types of sentences more similar to those they are exposed to in everyday life, Kidd et al., (2007) developed a more realistic picture of preschool children's comprehension abilities. The authors used presentational-style right-branching relative clauses, which are the most common type of relative clause produced by children (Diessel & Tomasello, 2005) (see examples (6)a. and b. below, from Kidd et al. 2007, p. 869). Presentational style refers to sentences where the subject of the matrix clause is a demonstrative pronoun (*this*), and the matrix verb is a copula. Right-branching relative clauses are those that modify the object of the matrix verb, rather than the subject.

(6) a. This is the boy that the girl teased at school yesterday. (Object relative clause)

b. This is the boy that teased the girl at school yesterday. (Subject relative clause)

With more appropriate sentence types in hand, Kidd et al. (2007) manipulated the animacy of the head nouns ("the boy" in both (6)a. and b.). In a sentence imitation task they found that for both English and German children, processing of object relatives with inanimate heads, (e.g., "Here is the food that the cat ate in the kitchen today"), was as easy as processing of subject relatives with animate heads. In the task, participants repeated a sentence immediately after the experimenter. Processing difficulty was measured as a function of repetition errors. A similar pattern of results has been found with a referential selection task, in which children selected toy objects being referred to using information from the relative clause (Brandt, Kidd, Lieven & Tomasello, 2009). For example, after watching actions acted out with toys, the experimenter asked the child "Can you give me the cake that the uncle stole?" and the participant would hand the experimenter one of two toy cakes. This method ensured that results were due to comprehension processing abilities, as sentence imitation tasks can be affected by memory load as well as linguistic ability. Overall, results from these studies mirror the adult results reviewed above: When semantic animacy is controlled for, differences in ease or challenge of comprehending subject versus object relative clauses disappear.

Relative clauses are a complex sentence form in which semantic animacy affects ease of processing in both adults and children. While semantic animacy is an inherent property of the noun, context dependent information, such as visual context, has also been shown to have effects on language comprehension.

Visual Information Use in Language Comprehension

Constraint-based theories of language processing argue that non-linguistic cues are used during online interpretation at the same time as linguistic cues to inform language comprehension. Tanenhaus et al. (1995) were the first to show that visual context is used immediately by adults to constrain sentence comprehension. They presented temporarily ambiguous sentences ("Put the apple on the towel in the box") to participants looking at a display of objects and locations, displaying either a single possible referent (one apple on a towel) or two possible referents (both an apple on a towel and an apple on a napkin). In these types of temporarily ambiguous sentences, the modifier "on the towel" creates a garden-path, causing listeners to parse the sentence in error, and requiring them to reinterpret the sentence when they hear "in the box". In the single referent condition, participants mistakenly interpreted "on the towel" as a location rather than a modifier and looked from the apple to an empty towel before turning their gaze to the box. In the two-referent condition, participants correctly used "on the towel" to distinguish which apple needed to be moved to the box without looking towards the empty towel. This illustrated the immediate use of visual cues in the interpretation of sentences, so that the listeners would not initially mis-parse the sentence online. Other studies have replicated and extended these findings (Chambers, Tanenhaus & Magnuson, 2004; Ferreira et al., 2013; Spivey, Tanenhaus, Eberhard & Sedivy, 2002).

Children's ability to use visual information to constrain possible interpretations during language comprehension is less certain. Trueswell et al. (1999) replicated the Tanenhaus et al. (1995) study described above with five-year-old children. Recall that in the original study, the number of apples in a visual display affected adult processing of the sentence "Put the apple on the towel in the box". The children in the Trueswell et al. (1999) study, however, were not able to use the two-referent condition to aid in comprehension of the ambiguous sentence "Put the frog on the napkin in the box." Instead they looked towards the empty napkin in both the single referent and two-referent conditions, suggesting that they interpreted "napkin" as the goal location. That is, having another frog that was not on a napkin did not cue the children that "on the napkin" was a modifying phrase specifying *which* frog to move. Hurewitz, Brown-Schmidt, Thorpe, Gleitman and Trueswell (2000) and Snedeker and Trueswell (2004) found similar results, although eye-movements showed that five-year old children were perhaps becoming sensitive to the referential distinction. However, in these studies many of the children also *never* reached the correct final interpretation of the sentence stimuli. This suggests that perhaps the tasks were too demanding for five-year olds and that the sentence structure was too complex for the developmental level of the children to be able to integrate multiple types of information. As Kidd et al. (2007) had found, using developmentally appropriate stimuli can greatly affect children's performance on experimental tasks, and the conclusions we may draw from them. Weighall (2008) also found similar results while varying the mode of presentation to reduce pragmatic misinterpretations and premature interpretations. For example, instead of displaying one frog on a napkin, and another on a platform, they displayed the frogs on different coloured napkins, to better equate the visual saliency between two frogs. Kidd, Stewart and Serratrice (2011) used simpler, syntactically ambiguous sentences such as "Cut the cake with the candle" to investigate whether children would use visual scene when interpreting such sentences. The visual display included a cake, a cake with a candle, a candle, and a knife. They found that adults would simultaneously consider both a plausible interpretation (i.e., candle as a modifying noun phrase) and an implausible interpretation (i.e., candle as an instrument) and their final interpretation was generally in favour of the plausible events. Five-year-old children on the other hand, strongly favoured the implausible interpretation, despite looking the most at a more plausible tool, the knife. The children were only able to parse the sentences to a single, improbable interpretation, despite the visual scene allowing for two interpretations, the other being more probable. These studies suggest young children may not be able to use visual information to modify their interpretations during language comprehension.

In contrast to the above findings, there is also some evidence that children may be able to use visual context during language processing (Huang & Snedeker, 2008; Van Rij, Hollebrandse & Hendriks, 2016; Weighall & Altmann, 2011; Zhang & Knoeferle, 2012). In contrast to studies with preschool-aged children, Weighall and Altmann (2011) found that slightly older children (ages six to eight years) were able to use visual context when processing relative clauses. Their findings were mediated by verbal memory span and whether the relative clauses were right-branching or centrally-embedded. Children with higher memory spans were more likely to show integration of the visual context, suggesting that cognitive capacity underlies information integration. There was also a positive effect for visual integration for items containing right-branching relative clauses. In addition, Zhang and Knoeferle (2012) found that German five-year-old children did use visual depiction of an action to aid in comprehension of Object-Verb-Subject (OVS) sentences, although less quickly than adults in an eye-tracking study. Huang and Snedeker (2008) found that five year olds used visual scene information when interpreting scalar adjectives 'big' and 'tall'. Van Rij et al. (2016) found that incongruent visual scenes elicited ungrammatical interpretations of object pronouns in four to six year olds, suggesting they were relying more on non-linguistic information than linguistic information. Thus, it appears that the ability for children to use visual information during processing depends on many factors, including the task, task demands, the syntactic forms in question (e.g., garden path sentences such as in Trueswell et al., (1999) vs. OVS sentences in Zhang and Knoeferle (2012)), the type of visual cues being presented, and individual factors such as age and memory span.

Semantic vs. Pragmatic and Non-linguistic Information

Although semantic, pragmatic and non-linguistic sources of information are known to be taken into account during adult language processing, they are traditionally viewed as distinct types of information. However, the distinction between semantic information, pragmatic information and non-linguistic information in the language processor may not be as clear as once thought (Nieuwland & Van Berkum, 2006; Vogels, Krahmer & Maes, 2013). In one illustrative example, Nieuwland and Van Berkum (2006) had participants listen to five-sentence stories with semantically anomalous characters (Experiment 1). In these stories, inanimate characters were treated as animate, thereby violating processing assumptions. For example, they outlined a story about a yacht receiving psychotherapy over his fear of water. Using event related potentials, they found that as the stories progressed, the strength of the N400 component, which responds to semantically anomalous language, decreased, showing adaptation to the appearance of an inanimate agent. This shows the processor's ability to reconcile atypical, improbable input to broader context. In their second experiment, they took this adaptation a step further. They created more elaborate stories, for example a story about a peanut performing a dance and a song about his romantic relationship with an almond. They found that at the end of a story centered around an inanimate agent (a dancing peanut), the N400 was stronger when the character was described with a semantically fitting yet pragmatically incongruent characteristic ("salted") than when the character was described with a pragmatically fitting yet semantically incongruent characteristic ("in love"). That is, listeners were more accepting of a pragmatically appropriate descriptor than a semantically appropriate descriptor. The context provided by the preceding story overruled semantic relatedness.

Vogels et al. (2013) varied the manner in which geometric shapes moved across a screen. When participants perceived the movement of the shapes to be more volitional, they used animate pronouns during retelling. When participants perceived the movement of the shapes to be less volitional, they used inanimate pronouns. Through this task, Vogels et al. (2013) demonstrated that non-linguistic information, visual context, also affects use of semantic animacy in language production. These studies illustrate how our use of linguistic knowledge can be affected by the context of input; they may not be completely distinct forms of information. Instead, our lexical knowledge may be considered a default expression of how we perceive the world is – if, for example, the current world in question appears to be a fictional world, our semantic representations may adapt to that context.

Nieuwland and Van Berkum (2006), when debriefing with participants, found that while listening to the test stories, they visualized the nouns (such as the peanut) in anthropomorphic ways, such as cartoon-like, with faces and limbs. Thus, in the current study we investigate whether providing such cues visually modulate the listener's comprehension.

In summary, semantic animacy information is used by both adults and children as a guide to interpretation when processing object relative clauses. However, it is unclear if and in what situations children can use visual information similarly to adults to constrain language comprehension. Lastly, there appears to be a flexible boundary between semantic information, pragmatic information, and non-linguistic information in adult language processing, yet the extent of this flexibility is unclear, especially in children.

This Study

This study examined how children resolve semantically implausible sentences with the use of visual cues. We used presentational style right-branching object relative clauses to investigate semantic and visual cues used in the language processing of adults and four- to five-and-a-half year old children. Using this form for experimental stimuli allowed us increased confidence that we would find effects of semantic animacy on processing at the outset of the study, as animacy information is a reliable cue during typical comprehension of such forms. We chose presentational style, right-branching object relative clauses for the experimental stimuli in this study due to the developmental appropriateness of the sentence frame (Diessel & Tomasello, 2005; Kidd & Bavin, 2002; Kidd et al., 2007). This allowed clearer interpretation of their ability to integrate visual cues with linguistic cues. Crucially, object relative clauses separate the grammatical subject and the first-mentioned noun, which subject relative clauses do not. For example, in the sentence "This is the boy that the girl teased at school yesterday", while "the girl" remains the grammatical subject, "the boy" has become the first-mentioned noun. If adults and children are cued only by subjecthood, one would find a clear subject preference, regardless of semantic animacy. However, Järvikivi, Van Gompel, Hyönä and Bertram (2005) found that adults use both subjecthood and order-of-mention as separate cues that affect the antecedent chosen during pronoun resolution. Thus, these cues may pull antecedent preference between the subject and the first-mentioned noun. By choosing a form that separates these two constraints, we increased the possibilities of uncovering an effect of semantic animacy, which may have otherwise been overshadowed by both subject preferences and first-mention preferences pointing toward the same noun.

To investigate the relationship between semantic and visual cues in language processing, we manipulated noun animacy. For example, in the sentence "There is the snake that the TV phoned at school", the TV, an inanimate object, takes on an animate, agentive role. In the absence of context suggesting otherwise, this sentence is both pragmatically and semantically anomalous (it is counter to both general world knowledge of what TVs and snakes do in real life and lexical knowledge of how the inanimate word *TV* should be used). These sentences were followed by sentences containing an ambiguous pronoun, *he* (e.g., "He was excited to go to the movie that night"). In general, ambiguous pronouns are resolved to the subject of the preceding sentence, or the first-mentioned noun of the preceding sentence (Järvikivi et al., 2005). However, in previous studies, all nouns followed appropriate rules of semantic animacy, and it is not known if such information is used in the same way with nouns that deviate in semantic animacy. In this study, we expected that pronoun resolution would show more preference to "the snake" as the pronoun's referent, due to it being the only semantically animate noun. Pronoun resolution preference would be demonstrated online, by looks to images during pronoun presentation, as well as offline, through verbal indications of pronoun antecedent. If accompanying images of the nouns included cartoon facial features on the image of the TV, we expected that the preference for "the TV", would increase, as it may seem more animate with a face. In cases where both nouns were semantically inanimate (e.g., "There is the couch that the TV phoned at school"), we expected pronoun resolution to follow similar patterns as if both nouns were animate (e.g., "There is the bunny that the TV phoned at school"), as semantic and visual animacy would no longer increase the prominence of one noun over another.

How would children perform with such anomalous sentences? Preschool-aged children are exposed to many situations through television and story books where inanimate objects are main characters of the plot. These characters have feelings, volitions, and the ability to act like humans, though they are everyday objects such as toasters (for example, the 1987 film *The Brave Little Toaster)* and vehicles (for example, the 2006 film *Cars)*. The average adult, although capable of imagination and understanding of fictitious worlds, is not exposed proportionally to the same quantity of cartoon situations in their day-to-day life in comparison to the preschooler. The preschooler's familiarity with such situations may make children more able to adapt to improbable language. Conversely, the mixed evidence suggests that children cannot always use visual cues to aid in comprehension, and so there may be no effects of facial features with children.

Research Questions and Hypotheses:

(1) How do children resolve ambiguous pronouns following semantically unusual sentences in comparison to adults? We hypothesized that when one noun was semantically inanimate, both children and adults would be more likely to choose the animate noun as the

pronoun antecedent regardless of grammatical role. This would demonstrate that semantic animacy modulates pronoun resolution in both children and adults. We expected this to be the case for both online, eye-tracking data and offline, verbal responses to comprehension questions.

(2) How do children use visual cues of animacy during ambiguous pronoun resolution in comparison to adults? We hypothesized that when presented with images of inanimate nouns that contained cartoon-like facial features, both children and adults would be more likely to consider those nouns as pronoun referents than when the images of inanimate nouns did not have facial features. Therefore we expected to see an effect of visual animacy on processing for both children and adults, as facial features would provide the necessary context needed for participants to consider inanimate nouns as appropriate pronoun antecedents. This would demonstrate that visual context modulates the use of semantic information in both adults and children. Again, we expected this to be the case for both online, eye-tracking data and offline, verbal responses to comprehension questions.

Method

In this study, participants listened to sentence pairs while their eye fixations to images on a Visual World scene were tracked. The Visual World scene consisted of an array of three images depicting the nouns included in the auditory stimuli (subject, object, location). The experiment contained two within-subject levels of manipulation: semantic animacy of the nouns and visual animacy of the inanimate noun images. Thus, nouns were either animate animals, or inanimate objects. Images of the inanimate objects were considered “visually animate” when they included facial features (otherwise were considered “visually inanimate”). The first sentence of each sentence pair always consisted of an object relative clause containing the manipulated nouns. The second sentence always began with the ambiguous pronoun *he* and contained generic information that could refer to either noun (Table 1). We measured proportions of looking times to the images after the onset of the pronoun. Prior research shows that children and adults reliably look at the image of a noun when they hear a pronoun that they believe refers to the noun (Arnold, Eisenband, Brown-Schmidt & Trueswell, 2000; Pyykkönen, Matthews & Järvikivi, 2010; Järvikivi et al., 2005; Song & Fisher, 2007). For example, after having heard "There is the couch that the TV phoned at school", if a child looks primarily at the TV after the onset of "He ...." we can assume that he or she has interpreted the TV as the pronoun's referent. After each item, the experimenter asked a simple comprehension question regarding the referent of the pronoun to confirm participants' final interpretation of the sentence and to compare to results of the eye gaze data.

Table 1

Sentence Manipulations

	Head Noun (Object) Animacy	**Second Noun (Subject) Animacy**	**Sentence**
Sentence 1:	<u>Inanimate</u>	<u>Inanimate</u>	There is the <u>couch</u> that the <u>TV</u> phoned at school.
		Animate	There is the <u>couch</u> that the *bunny* phoned at school.
	Animate	<u>Inanimate</u>	There is the *snake* that the <u>TV</u> phoned at school.
		Animate	There is the *snake* that the *bunny* phoned at school.
Sentence 2:			He was excited to go to the movie that night.

Participants

Two age groups of participants – adults and young children – participated in the experiment. Forty-three adult undergraduate university students were recruited through a participant database through the University of Alberta Linguistics Department and received course credit for participation. All adult participants completed the experiment at the Centre for Comparative Psycholinguistics lab. Five of the adult participants were excluded from the analysis due to bilingual language background, and two of the participants were excluded due to technical difficulty with maintaining eye-tracking. After exclusions, there were 36 participants in the adult group (ages ranged from 18 to 52 years old, mean = 21.16; 30 females). All included participants were monolingual English speakers, had normal or corrected to normal vision, and normal hearing. Monolingual status was determined through self-report (see attached document in Appendix A)– participants who reported a first language other than English, or rated themselves as fluent in a language other than English were excluded. Typical language development was determined through self-report.

Forty-eight children aged 4;0 – 5;5 years old participated in this study. All included participants were monolingual English speakers, had normal or corrected to normal vision, normal hearing, and typical language development as determined through parental report (see attached document in Appendix B). Children reported as having another first language other than English, or were rated by their parents as presenting any fluency in a second language were excluded. Children who were reported as being exposed to another language but not yet presenting with any fluency in the second language were included in the study.

Typical language development of the child participants was determined primarily by parental report. In addition, each child completed the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals – Preschool 2 (CELF-P2) (Wiig, Secord & Semel, 2004) as an additional screen of language abilities after the experimental protocol. Sentence repetition tasks have been shown to be strongly sensitive to language impairment (Ramsden, Botting & Faragher, 2001). Any participant whose scaled score was more than one standard deviation below the population mean (below a scaled score of 7) for their age would be excluded. All child participants' scores were above -1 standard deviation (scores ranged from 7 to 18, mean = 12.17). One participant would not participate in the sentence repetition task due to attention fatigue and therefore her score was not obtained. No parents reported concerns with language development.

Participants were also excluded from the study if they did not correctly answer at least 14/20 comprehension questions asked throughout the study. Children who could not reach this criterion were considered unable to reliably comprehend or attend to the task. Only comprehension questions from filler items (described below) were used to determine inclusion.

We recruited the children through community preschools and daycares, as well as through the Child and Adolescent Research Group database at the University of Alberta and through word-of-mouth in the community. Participants were run in quiet rooms/areas within their preschools and daycares, or were brought into the Language Development and Disorders Lab by their parents. Child participants received a t-shirt for participation. Four of the child participants were excluded from the study due to attention/comprehension difficulties during the study (as measured by incorrectly answering more than 30% (6/20) of the filler comprehension questions during the experiment). Three of the participants chose to discontinue the experiment before completion, two participants were outside the specified age range, one child was excluded due to technical difficulty maintaining eye-tracking, one child was excluded due to consistently not looking at the screen during the experiment, and one child was excluded due to bilingual language background. After exclusions, there were 36 participants (22 girls) in the child participant age group, ranging in age from 4;0 to 5;3 years old (*M* = 4;6).

Materials

Sentence pairs. We created twenty sentence pair test items. The first sentence started with a demonstrative pronoun (*here, there, this* and *that*) and contained a presentational style right-branching object relative clause. As previously noted, presentational style right-branching object relative clauses refer to sentences where the subject of the matrix clause is a demonstrative pronoun, the matrix verb is a copula, and the relative clause modifies the object of the matrix verb, as illustrated in example (7).

(7) a. This is the boy that the girl teased at school yesterday. (Kidd et al., 2007)

b. There is the couch that the TV phoned at school. (This study)

Both the subject and object noun of the relative clause were manipulated for semantic animacy (animate animal vs. inanimate object). There were therefore four versions of each test item. An example of the sentence versions can be seen in Table 1, listed again below, and the full list of test sentences can be found in Appendix C. Subject and object nouns used in this study had a minimum frequency of 50 occurrences per million words for the age range of 48 to 66 months in the ChildFreq database (Bååth, 2010) to increase assurance of familiarity among the child participants. Verbs were borrowed from Pyykkönen et al. (2010), as these verbs showed a preference for the subject as pronoun antecedent in that study. Thus, the default preference for pronoun referent should be the subject of the relative clause for all test items. Each sentence ended with a prepositional phrase referring to a location to pull eye-gaze away from the images of the subject and object before the pronoun was presented.

Table 1

Sentence Manipulations

	Head Noun (Object) Animacy	**Second Noun (Subject) Animacy**	**Sentence**
Sentence 1:	Inanimate	Inanimate	There is the couch that the TV phoned at school.
		Animate	There is the couch that the *bunny* phoned at school.
	Animate	Inanimate	There is the *snake* that the TV phoned at school.
		Animate	There is the *snake* that the *bunny* phoned at school.
Sentence 2:			He was excited to go to the movie that night.

The second sentence of each pair began with the masculine third person singular pronoun, *he*, to maintain maximal ambiguity to the referent. The sentence following the pronoun also aimed to remain ambiguous to the referent by describing a state or action potentially appropriate to either sentence character. Therefore pronoun resolution could not have been affected by the content of sentence 2. It is pertinent to recognize that all sentence items are pragmatically unusual. While it is improbable that a TV would phone a couch, it is also improbable that a bunny would phone a snake. Consistency in probability across sentence types contributes to confidence that significant results are attributable to the semantic and visual manipulations of the experimental items only.

Twenty filler items were also presented during the experiment. The filler items were similar to the test items in that they consisted of two sentences, yet they did not include pronominal reference or relative clauses. The first sentence was a simple active declarative sentence ending with a location preposition similar to the test items. All subjects and objects in the filler sentences were animate, either human or animal. The second sentence directly referred to the subject or object of the previous sentence. For example, “The boy swam with the man at the lake. The man found it very cold”. A list of all filler sentences can be found in Appendix D.

Test and filler items were recorded by an adult female North American English speaker in a sound attenuated booth at a sampling frequency of 44.1 kHz over two recording sessions. The speaker maintained a broad focus prosodic contour for each item (used when answering a “what happened?” question) and was coached to speak as if speaking to a preschool aged child. We segmented each individual sentence into a separate audio file to standardize pause length between sentence 1 and sentence 2. Audio clip onsets were cropped to the first indication of articulation on the audio waveform. Audio clip offsets were cropped to the end a phonemic sound. For example, sentences ending in a nasal sound (e.g., /m/) were cropped to the end of periodicity in the waveform, and sentences ending in a fricative (e.g., /s/) were cropped to a sharp decrease in aperiodic noise amplitude. Cropping the end of sentence 1 erred on the side of over-cropping, to avoid silence at the end of the audio clip. A silent pause of 800 ms between sentences 1 and 2 was programmed into the experiment.

Visual Stimuli. Images of the two characters (couch/snake and TV/bunny) and the location (school) from the first sentence were displayed on screen 1000 ms before the onset of audio and for the duration of each test item. The visual display was removed during the comprehension question. We collected cartoon images from two open-licensed clip art websites (Clikr.com, Openclipart.org) as well as from Shutterstock.com through a licensing agreement.

We manipulated visual animacy of the inanimate noun images. From the original picture of the object, we created a second picture with cartoon eyes and a mouth superimposed with online image editors (ImageBot, Online Image Editor) as well as with Adobe Photoshop (Version 14.0). In one half of the trials, participants were shown the original image (visually inanimate:

no face). In the other half of the trials they saw the manipulated image (visually animate: with a face). Visual animacy was blocked within the experiment and block order was counterbalanced between participants. That is, images with faces were included in either the first half of the experiment or the second half of the experiment.

Figure 1 is an example of the Visual World display where both objects have been manipulated to be visually animate. The three images were placed in counterbalanced orders in the top centre, bottom left corner, and bottom right corner of the screen. Each image was placed inside a 370 pixels wide by 330 pixels high interest area. From centre of the screen, (0,0) pixels, the centre of the top centre interest area was located at (0, 347) pixels. The centre of the bottom left interest area was located at (-327, -347) pixels, and the centre of the bottom right interest area was located at (327, -347) pixels. Looks to interest areas were recorded during the experiment.

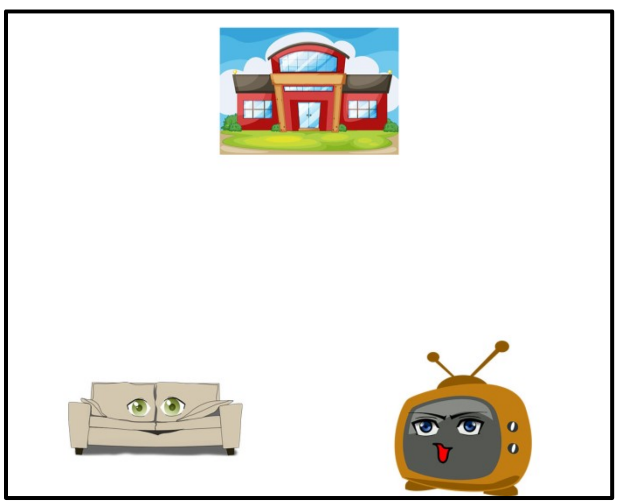

Figure 1: Visual World display of test item "There is the couch that the TV phoned at school. He was excited to go to the movie that night." with images manipulated to add visual animacy.

Procedure

In total, there were 20 sentence pairs, each with four versions to manipulate semantic animacy (inanimate-inanimate, inanimate-animate, animate-inanimate, and animate-animate, as per Table 1). Each participant heard only one version of each sentence pair, for a total of twenty test items. The experiment was divided into two ten-item blocks to manipulate visual animacy (counterbalanced across participants). That is, participants saw visually animate images for either the first or second half of the experiment. Within each block, we presented all four sentence types of semantic animacy manipulations to the participants. Therefore, it was necessary to have an unequal number of sentence types occurring in each block. For example, within a single block, one could have two inanimate-inanimate items, two inanimate-animate items, three animate-inanimate items, and three animate-animate items, for a total of ten test items. Then in the second block, the same participant would have three inanimate-inanimate items, three inanimate-animate items, two animate-inanimate items, and two animate-animate items to cover the remaining ten test items. To minimize any uncontrolled effects of these unequal distributions, we created six lists to counterbalance the sentence versions across lists. Twenty filler items were included (ten in each block), for a total of forty sentence pairs. Across the 36 participants in each age group, six participants were assigned to each list, and visual animacy was counterbalanced within those participants (i.e., three participants saw visually inanimate images first, and three participants saw visually animate images first). Within each block, the order of the items (both test and filler items) were presented in a random order, with the exception of the very first item of the experiment, which was always a filler item to allow as a practice trial.

In the experiment, the child participants were seated in front of a computer screen. Monitor size was 1280X1024 pixels. Eye movements were tracked with an arm-mounted SR Research Eye Link 1000 Plus head-free eye-tracker, sampling at 500 Hz. Participants' right eyes were always tracked. The experiment was presented with SR Research's Experiment Builder software (Version 1.10.1025). Before the experiment began, the eye-tracker was calibrated with a nine-point calibration display, and a comfortable listening volume was established for the participant. The experimenter instructed the child participants that they were going to hear some silly stories while they looked at pictures on the screen, and answer some questions. To start each trial, participants looked at a cartoon star as a fixation point presented in the middle of the screen. The experimenter triggered the visual world display manually when the child fixated the star. The visual display was presented for 1000ms, after which the corresponding sentence pair was played over speakers. After each trial, the experimenter verbally asked the participants a comprehension question formulated from the second sentence, by simply replacing the pronoun *he* with *who*. For example, for the sentence pair “There is the couch that the TV phoned at school. He was excited to go to the movie that night”, the experimenter then asked, “Who was excited to go to the movie that night?” Filler comprehension questions were designed in the same way. The comprehension questions served two purposes: in the case of the test items, it allowed us to gather offline information on the participants’ final interpretation of the pronoun. In the case of the filler items, comprehension questions had specific correct/incorrect responses, which were tallied to guarantee a minimum level of attention and comprehension as mentioned above in the exclusion criteria. When participants were unsure how to answer the questions, or looked to the experimenter for help, they were encouraged to do the best they could, and guess if they needed to. If they were still hesitant to guess, or if they had not listened to a particular trial, the experimenter suggested that “I don’t know” answers were accepted. Answers approximating the vocabulary used in the trial, such as “boot” for “shoe” were accepted as referring to the intended noun when referent was clear. Ambiguous answers such as “the guy”, random answers, and “I don’t know” answers were coded as referring to neither the subject nor the object of sentence 1. Specific answers were written down by the experimenter and coded as referring to the subject, object, or neither with a key press during the experiment.

Between block 1 and 2, an optional movement break was offered to participants. After the experiment, the child participants were given the CELF-P2 Recalling Sentences subtest.

The procedure for adult participants was similar to the child participants, with some differences. While 17 of the adult participants were run on the same eye-tracker as the child participants, 19 of the adult participants were run on a table-mounted SR Research Eye Link

1000 eye-tracker with a headrest, also sampling at 500 Hz. Screen size for was 1920X1080 pixels. Adult participants were told they would be completing the same experiment as children, and that they would listen to sentences while looking at pictures on a screen. However, instead of hearing the comprehension questions spoken by the experimenter, adult participants read the comprehension questions on the screen and answered aloud.

Results

Analysis Methods

In Visual World eye tracking studies, the measured dependent variable is the proportion of looks (here, transformed into logits, as explained below) to a target interest area time-locked to a specific point of interest in the auditory stimulus. In this particular study, we were interested in the listener's referential preference for either the subject or the object when hearing the pronoun "he". Thus, we were interested in the relative proportion of fixations to the pictures that represent the subject and object character. The independent variables were the sentence type (manipulations on semantic animacy) and image condition (manipulations on visual animacy).

Raw eye-tracking data tell the researcher where the participant's eyes were looking at a single moment every two milliseconds for the entire analysis time frame. Each trial therefore produced 500 data points per second, which were then aggregated to the proportion of looks to each interest area over a 20 millisecond "bin". For statistical analysis, the proportion of looks in each bin was then transformed into empirical logits (see Barr, 2008). Therefore, our dependent variable is the difference of looks to the subject interest area from the looks to the object interest area in logits. Positive values indicate a preference for the subject, while negative values indicate a preference for the object. Each logit within the analysis time frame is a separate data point. Data preparation was completed in the statistical processing software R (R Core Team, 2016) with the preprocessing package *VWPre* (Porretta, Kyröläinen, Van Rij & Järvikivi, 2016).

The data were divided into three analysis timeframes. The first time frame was between 200 ms post onset of the pronoun and 800 ms post onset. The 200 ms delay in time-frame 1 is to account for the minimum time needed to program and execute an eye-movement (Matin, Shao & Boff, 1993). Time-frame 2 was 800 ms to 1400 ms post onset, and time-frame 3 was 1400 to 2000 ms post onset (see Clackson, Felser & Clahsen, 2011; Järvikivi, Pyykkönen-Klauck, Schimke, Colonna & Hemforth, 2014).

To analyze this data, we used linear mixed effects (LME) modelling with the package *lme4* (Bates, Maechler, Bolker & Walker, 2015). LME is a method of linear regression analysis that allows one to take into consideration both fixed and random effects. Fixed effects are the factors in an experiment that are manipulated, and therefore are repeatable. Random effects, such as participants and linguistic items, are non-repeatable factors that have been sampled from a larger population (of all people, or all sentences). These factors are modelled in LME as random variables with a mean of zero and unknown estimated variance (Baayen, 2008; pp 241-242). Accounting for the variance of participants and items is important for being able to generalize to the larger population of both people and languages (Baayen, 2008; Baayen, Davidson & Bates, 2008; Jaeger, 2008). Traditionally, other methods (e.g. quasi-F, Latin-square) have been used to account for random effects when analyzing data with ANOVA. LME is chosen as an analysis method over ANOVA because it is a more powerful analysis method than ANOVA as it requires no aggregation over participants and items, allows for testing continuous and categorical variables in a single model, and because it can better account for crossed by-item and by-participant random effects.

At any point in time, participants in a Visual World experiment can look at one of a discrete number of interest areas at a time. Therefore, eye-tracking data are purely categorical. Jaeger (2008) argued that for categorical data such as eye-tracking data, ANOVA methods, even when used with arcsine-square-root transformed proportions of categorical outcomes, can produce incorrectly significant or null results beyond typical Type I and Type II errors rates.

Using categorical eye-tracking data as if it were scalar violates all assumptions of ANOVA. Instead, Jaeger (2008) suggests using mixed logit models, a type of Generalized Linear Mixed Modelling that transforms the categorical proportional data from eye-tracking into a logit scale. The logit scale[1] has a mean of zero, and is completely unbounded. This transformation is necessary because when categorical data are bounded between 0 and 1 in a proportional scale, changes near chance (0.5) reflect smaller changes in predictors than changes near the bounds (0 and 1). By unbounding the data with a logit scale, the effects of predictors become linear and thus analyzable. Other benefits of using an LME model over ANOVA include that LME can include both categorical and continuous predictors, provides more information of the directionality and size of the effect, does not necessitate the assumption of homogeneity of variances, is more powerful, and is robust against missing data (Baayen et al., 2008; Jaeger, 2008; Quené & Van den Bergh, 2008). Importantly, LMEs do away with the traditional averaging over subjects and items and thus result in significantly less loss of variance and, consequently, more power.

Additionally to being a more effective method of analysis for categorical data than ANOVA, LME modelling accounts for random effects more efficiently than other methods. Baayen (2008; pp. 260-269) and Baayen et al. (2008) compared LME to quasi-F and Latin Square designs. While quasi-F shows similar power and Type I error rate for calculating random effects, LME is less vulnerable to missing data points. In comparison to Latin Square designs, LME is more powerful and less conservative. Also, LME allows covariates to be taken into account and gives more insight into participant and item effects which neither quasi-F nor Latin Square can do.

To summarize, the data from this experiment are categorical, as at any individual time participants can look at only one of the three interest areas. Analysis of categorical data with

---

1 $\ln \frac{p}{1-p} = \text{logit } p$ (Jaeger, 2008: page 437)

ANOVA, although common, violates assumptions and produces unreliable results. Additionally, LME allows us to include continuous covariates of age, trial and Recalling Sentences subtest scores to our models. The models are robust against missing data, an important benefit for eye-tracking data, especially with children. And, importantly, it also allows us to take into account the random effects of participants and items, making results more generalizable and our models more powerful.

In addition to the eye-tracking data, this experiment also generated data from the participants' verbal answers to the comprehension questions, coded as either referring to the object of the preceding sentence, the subject of the preceding sentence, or 'other'. 'Other' responses were discarded for statistical analysis. We analyzed the resulting binomial data with generalized linear mixed models using *lme4* (Bates et al., 2015), using verbal answer codes "subject" or "object" as the dependent variable.

The analysis results will be reviewed in the following way: All child data will be presented, followed by all adult data. First, we present results of an overall analysis of the eye-tracking data to illustrate the effects semantic animacy had on pronoun resolution. Then analyses of each individual sentence type will be presented to illustrate effects of visual animacy. Offline verbal response data will then be presented in the same manner.

Children: Eye-tracking Data Results

The research questions asked how children use semantic animacy and visual cues of animacy during language processing in comparison to adults. The primary analysis examined whether children processed pronouns differently as a function of the semantic animacy of the subject and object nouns of the preceding relative clause, and whether the addition of facial features to semantically inanimate objects affected language processing. All models included subject and item intercepts and by-subject and by-item random slopes for sentence type and image condition. Covariate predictors of age (in months), age group ("48-55 months", or "57-63 months"), raw Recalling Sentences score, scaled Recalling Sentences score, and trial order were included in the models, and removed if they did not add to the fit. Trial order was the only covariate that remained in the final models.

Overall analyses. To target the question of whether semantic animacy affected pronoun resolution, we fit a model on the data that included sentence type and image condition as fixed effect predictors. Sentence type contained four levels: animate-animate (AN_AN), inanimate-inanimate (IN_IN), inanimate-animate (IN_AN), and animate-inanimate (AN_IN), where the animacy is coded by order of mention – that is, the first noun refers to object, second noun refers to subject. For example, "There is the snake that the TV phoned at school" is an AN_IN sentence. Image condition contained two levels: NoFace, in which inanimate objects did not have facial features, and Face, in which eyes and mouths had been added to the images. It is important to remember here that image condition coded for the block type, not just whether the images were semantically animate or not. That is, the AN_AN sentences were still coded for Face/NoFace condition, despite the fact that the images for these sentences were by default visually animate by virtue of containing only animate nouns in them. However, an effect of surrounding context may or may not still have an effect on the AN_AN sentences, so it was important to do an analysis in this way.

We used AN_AN NoFace items as the reference condition (i.e., the intercept) which all conditions were compared against. Results are summarized in Table 2, with statistically significant findings highlighted in bold font. Recall that because the dependent variable is a difference score between looks to subject and object, coefficients of positive value refer to more looks to the subject image, whereas coefficients of negative value refer to more looks to the object image. For example, if a predictor had a negative coefficient, there would be significantly more looks to the object noun in that situation, in comparison to the reference situation, AN_AN NoFace. The reference, AN_AN NoFace sentences, showed a statistically significant object preference in the first time frame (200-800 ms). That is, in "There is the snake that the bunny phoned at school", the children initially looked longer at "the snake", the first-mentioned

Table 2

Child eye-tracking results: Mixed-effects modelling of difference of logit looks to subject noun vs. object noun for three analysis time frames.

Predictor	**Time-frame**	**ß**	**SE**	**t**	**p**
Intercept	200-800 ms	-1.456	0.573	-2.540	**0.015***
	800-1400 ms	1.199	0.970	1.236	0.227
	1400-2000 ms	-0.473	0.575	-0.824	0.415
IN_IN	200-800 ms	0.803	0.591	1.360	0.181
	800-1400 ms	0.615	0.981	0.627	0.536
	1400-2000 ms	1.021	0.853	1.196	0.239
IN_AN	200-800 ms	1.646	0.744	2.212	**0.032***
	800-1400 ms	0.943	1.026	0.919	0.364
	1400-2000 ms	1.449	1.165	1.244	0.223
AN_IN	200-800 ms	0.027	0.699	0.038	0.970
	800-1400 ms	-1.705	1.136	-1.501	0.142
	1400-2000 ms	-1.169	0.703	-1.663	0.104
Face	200-800 ms	0.271	0.493	0.548	0.587
	800-1400 ms	0.780	0.502	1.555	0.127
	1400-2000 ms	0.924	0.510	1.810	0.077
Trial order	200-800 ms	0.041	0.005	7.781	**1.09e-14*****
	800-1400 ms	-0.050	0.006	-8.832	**< 2e-16 *****
	1400-2000 ms	-0.005	0.006	-0.871	0.384
IN_IN:Face	200-800 ms	-0.933	0.162	-5.764	**8.33e-09 *****
	800-1400 ms	-1.231	0.168	-7.345	**2.13e-13 *****
	1400-2000 ms	-0.938	0.170	-5.502	**3.81e-08 *****
IN_AN:Face	200-800 ms	-1.655	0.152	-10.917	**< 2e-16 *****
	800-1400 ms	-2.203	0.157	-14.020	**< 2e-16 *****
	1400-2000 ms	-2.035	0.160	-12.731	**< 2e-16 *****
AN_IN:Face	200-800 ms	0.346	0.162	2.139	**0.033 ***
	800-1400 ms	1.579	0.168	9.427	**< 2e-16 *****
	1400-2000 ms	1.789	0.170	10.511	**< 2e-16 *****

p < 0.05 *. p< 0.01 **. p <0.0001 ***

noun, than "the bunny", the subject of the sentence. In the latter two time frames, there were no significant differences in proportion of looks to the two animate nouns.

As illustrated in Table 2, there were a number of significant effects. Of primary interest is the significant difference between IN_AN sentences and AN_AN sentences (the reference) observed during time window 1 (200-800ms), showing an overall relatively larger subject preference (i.e., there were more looks to the subject image than the object image) when the object entity was inanimate than when it was animate. Using the example sentences, this means that "There is the couch that the bunny phoned at school" elicited more of a preference for the subject, *bunny*, than "There is the snake that the bunny phoned at school". In the first sentence, the object, "couch", is semantically inanimate, and therefore less suitable as a pronoun referent. This result illustrates that semantic animacy did affect children's pronoun resolution. This result of semantic animacy was necessary to establish before analyzing the effects visual animacy would have on inanimate nouns.

The table entries labeled IN_AN:Face, IN_IN:Face, and AN_IN:Face indicate that there were significant interactions between sentence type and image condition. For IN_IN and IN_AN sentence types, when faces were added, there were significantly more looks to the semantically inanimate, but visually "animate" object in all time frames in comparison to AN_AN sentences. The AN_IN sentences showed a stronger increase in looks to the semantically inanimate subject with faces added in all time frames in comparison to the reference condition. Lastly, other significant effects included trial order in the first two time-frames, which will be discussed in later sections.

Results from Table 2 are illustrated in Figure 2 (the same figure with error bars included can be found in Appendix E). Effects of semantic animacy discussed above are illustrated by comparing the IN_AN data (yellow) to the AN_AN data (blue) in the NoFace block. From 200 to 800 ms, IN_AN sentences show more looks to the subject, while AN_AN sentences show slightly more looks to the object. Notice also that data for AN_IN (green) appear different from the AN_AN (blue) data in the NoFace block. This trend, however, is not statistically significant. Interactions between sentence type and image condition are observed by looking at trends across the figures for all sentence types.

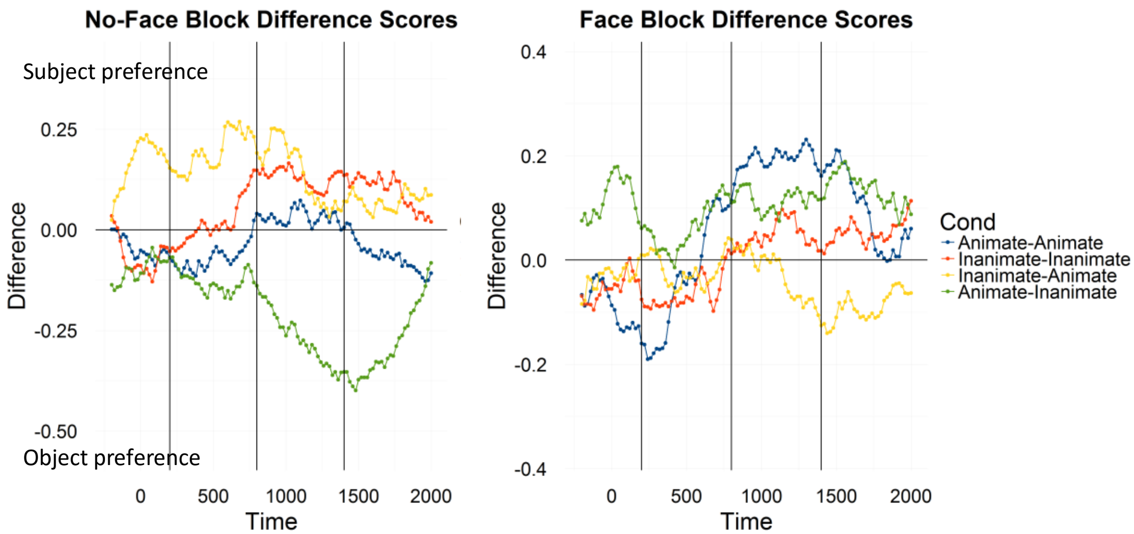

Figure 2. Child eye-tracking data by sentence type and image condition. Difference scores = looks to subject interest areas subtract looks to object interest areas. Vertical lines indicate 200, 800 and 1400 ms respectively.

Table 3

Child eye-tracking results: Mixed-effects modelling for each sentence type.

Sentence Type	**Predictor**	**Time-frame**	**ß**	**SE**	**t**	**p**
Animate – animate	Intercept	200-800 ms	-1.649	0.662	-2.490	**0.016** *
		800-1400 ms	3.226	0.948	3.403	**0.002** **
		1400-2000 ms	-1.285	0.884	-1.453	0.154
	Face	200-800 ms	0.201	0.924	0.217	0.830
		800-1400 ms	0.734	1.025	0.715	0.479
		1400-2000 ms	1.001	1.142	0.876	0.387
	Trial order	200-800 ms	0.054	0.014	3.974	**7.44e-05** ***
		800-1400 ms	-0.138	0.013	-10.404	**< 2e-16** ***
		1400-2000 ms	0.031	0.013	2.287	**0.022** *

Inanimate - inanimate	Intercept	200-800 ms	0.800	0.594	1.347	0.183
		800-1400 ms	2.269	0.805	2.820	**0.007** **
		1400-2000 ms	0.623	0.664	0.939	0.352
	Face	200-800 ms	-0.408	0.885	-0.462	0.647
		800-1400 ms	-0.462	1.063	-0.434	0.667
		1400-2000 ms	-0.128	0.979	-0.131	0.897
	Trial order	200-800 ms	-0.028	0.011	-2.589	**0.010** **
		800-1400 ms	-0.067	0.011	-6.078	**1.48e-09** ***
		1400-2000 ms	-0.003	0.012	-0.294	0.768
Inanimate - animate	Intercept	200-800 ms	-0.523	0.592	-0.883	0.381
		800-1400 ms	2.797	0.725	3.856	**0.0005** ***
		1400-2000 ms	0.173	0.832	0.208	0.836
	Face	200-800 ms	-1.600	0.821	-1.947	0.058
		800-1400 ms	-1.322	0.760	-1.740	0.092
		1400-2000 ms	-1.081	0.840	-1.286	0.207
	Trial order	200-800 ms	0.080	0.010	7.794	**9.33e-15** ***
		800-1400 ms	-0.077	0.011	-7.019	**3.68e-12** ***
		1400-2000 ms	0.042	0.011	3.595	**0.0003** ***
Animate – inanimate	Intercept	200-800 ms	-2.209	0.650	-3.396	**0.001** **
		800-1400 ms	-2.401	0.718	-3.346	**0.002** **
		1400-2000 ms	0.218	0.789	0.276	0.784
	Face	200-800 ms	1.079	0.823	1.311	0.198
		800-1400 ms	2.383	0.970	2.457	**0.019** *
		1400-2000 ms	2.575	1.036	2.485	**0.017** *
	Trial order	200-800 ms	0.066	0.012	5.560	**3.29e-08** ***
		800-1400 ms	0.033	0.012	2.772	**0.006** **
		1400-2000 ms	-0.097	0.013	-7.484	**1.01e-13** ***

p < 0.05 *. p< 0.01 **. p <0.0001 ***

Individual sentence-type analyses. In order to investigate the effects of visual animacy on pronoun resolution for each sentence type, we modelled each sentence type separately and included visual animacy as the only fixed effect. That is, for each sentence type, we compared the visually animate data against the visually inanimate data. Results of the analyses are illustrated above in Table 3.

Table 3 indicates that there was a significant effect of visual animacy only for AN_IN sentences (e.g., “There is the snake that the TV phoned at school”) in time frames 2 and 3. This indicates that without faces added to the inanimate subject nouns, there was a preference for the animate object nouns. When faces were added to the subject pictures, participants preferred the subjects as pronoun referents. This suggests that adding facial features to the inanimate pictures allowed the children to consider the noun “TV” as an animate entity. This effect is shown in Figure 3. Additionally the effect of Trial illustrated in the initial model was maintained for every sentence type. There was no significant effect of visual animacy for the other sentence types.

In summary, the children preferred semantically animate nouns over semantically inanimate nouns within time frame 1 (200-800 ms) for IN_AN sentences only. They preferred visually animate nouns over visually inanimate nouns within time frames 2 and 3 (800-1400, 1400-2000 ms) for AN_IN sentences.

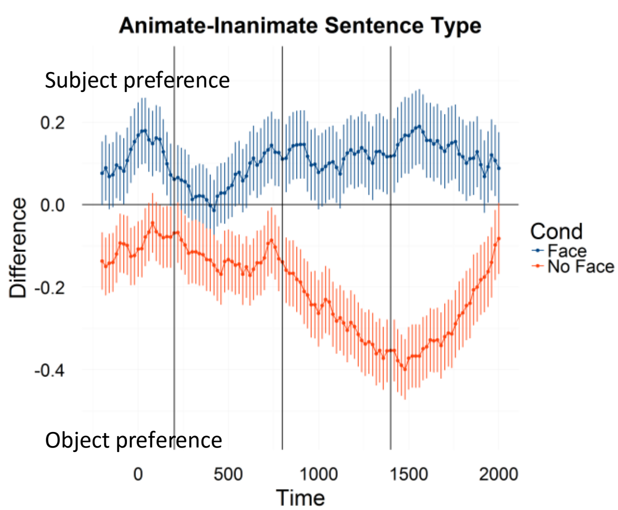

Figure 3. Child eye-tracking data for animate-inanimate sentence type. Difference scores = looks to subject interest areas subtract looks to object interest areas. Vertical lines indicate 200, 800 and 1400 ms respectively.

Children: Offline Verbal Response Results

Overall analysis. Analysis of the offline referent selection data followed the same process of investigating effects of first semantic animacy and then visual animacy. Models included only random effects for subject and item intercepts, due to lack of convergence when by-subject and by-item slopes were included in the models. Covariate predictors of age (in months), age group, raw Recalling Sentences score, scaled Recalling Sentences score, and trial order were all insignificant, and therefore removed from the final models.

Mirroring the eye-tracking analysis above, the overall model included sentence type (four levels: animate-animate (AN_AN), inanimate-inanimate (IN_IN), inanimate-animate (IN_AN), and animate-inanimate (AN_IN)) and image condition (two levels: NoFace, and Face). The reference condition was AN_AN sentence type in the NoFace block, similarly to the eye-tracking data. In contrast to the eye-tracking data where children initially preferred the object noun, and then looked equally to both nouns, children showed a slight, yet significant preference for the subject as antecedent in their verbal responses for AN_AN NoFace sentences (i.e., the intercept), as indicated by a significantly positive coefficient. Results are summarized in Table 4.

Table 4

Child verbal response results: Mixed-effects modelling of verbal response to comprehension questions indicating subject noun vs. object noun interpretation of ambiguous pronoun.

Predictor	**ß**	**SE**	**z-value**	**p**
Intercept	0.579	0.237	2.442	**0.015** *
IN_IN	-0.089	0.333	-0.267	0.790
IN_AN	1.593	0.423	3.770	**0.0002** ***
AN_IN	-1.353	0.328	-4.124	**3.73e-05** ***
Face	-0.378	0.320	-1.180	0.238
IN_IN:Face	0.256	0.459	0.557	0.578
IN_AN:Face	-0.888	0.534	-1.663	0.096
AN_IN:Face	1.154	0.455	2.540	**0.011** *

$p < 0.05$ *. $p < 0.01$ **. $p < 0.0001$ ***

In this data set, not only was the difference between AN_AN and IN_AN sentences again significant, the difference between AN_AN and AN_IN sentences was also significant. These differences are illustrated in Figure 4. In comparison to the AN_AN sentences (e.g., "There is the snake that the bunny…"), the IN_AN (No Face) sentences (e.g., "There is the couch that the bunny…") generated more responses indicating the subject (the bunny) as antecedent, as indicated by the positive sign of the estimate, whereas the AN_IN sentences (e.g., "There is the snake that the TV…") generated more responses indicating the object (the snake) as antecedent, as indicated by the negative sign. This strengthens the claim that the semantic animacy of the nouns affected the resolution of pronouns. Children were more likely to consider animate nouns as pronoun referents.

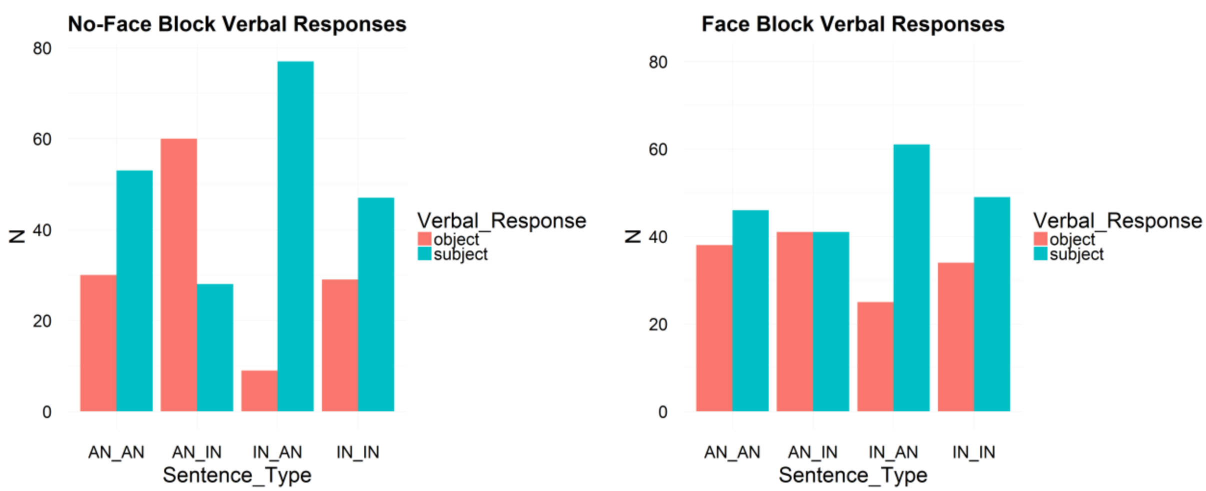

Figure 4: Child verbal response data by sentence type and image condition. N = number of verbal responses indicating pronoun referent.

Similarly to the eye-tracking data, there was also a significant interaction between sentence type and image condition, as seen in Table 4 (AN_IN:Face). For the AN_AN sentences, object responses increased when faces were added (though not significant as a main effect), while for the AN_IN sentences, subject responses increased when faces were added. To investigate whether or not adding facial features did affect pronoun resolution to some degree as main effects for the verbal response data, models of each sentence type were again created. Results are shown in Table 5.

Table 5

Child verbal response results: Mixed-effects modelling for individual sentence types.

Sentence Type	**Predictor**	**ß**	**SE**	**z-value**	**p**
Animate - animate	Intercept	0.594	0.274	2.166	**0.030** *
	Face	-0.384	0.337	-1.141	0.254
Inanimate - inanimate	Intercept	0.487	0.241	2.018	**0.044** *
	Face	-0.118	0.327	-0.360	0.719
Inanimate - animate	Intercept	2.833	0.603	4.698	**2.63e-06** ***
	Face	-1.584	0.525	-3.020	**0.003** **
Animate - inanimate	Intercept	-0.874	0.284	-3.075	**0.002** **
	Face	0.876	0.349	2.507	**0.012** *

$p < 0.05$ *. $p < 0.01$ **. $p < 0.0001$ ***

Individual sentence-type analyses. For both IN_AN (e.g., "… couch that the bunny…") and AN_IN (e.g., "…snake that the TV…") sentence types, adding faces did significantly affect pronoun resolution. The effects are illustrated in Figures 5 and 6. For IN_AN sentences (Figure 5), the subject preference seen without the facial features added is very strong, and while maintained, decreases in strength significantly when faces were added, indicating that when the object (the couch) has animate visual characteristics, it is more likely to be considered as a pronoun referent. For AN_IN sentences (Figure 6), there is a strong object preference when no faces are added, indicating that the animate noun, that is, the syntactic object (the snake), is considered more likely to refer to the pronoun than the inanimate noun (the TV). When facial features are added, the object preference disappears, and both nouns are equally considered for pronoun reference.

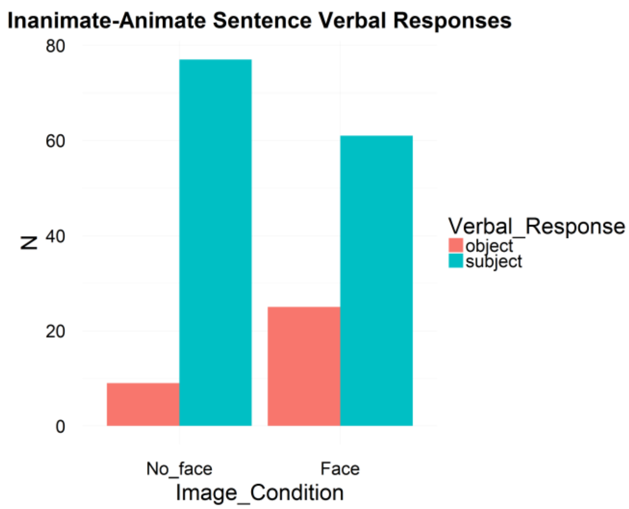

Figure 5: Child verbal response data by image condition for Inanimate-Animate sentences. N = number of verbal responses indicating pronoun referent.

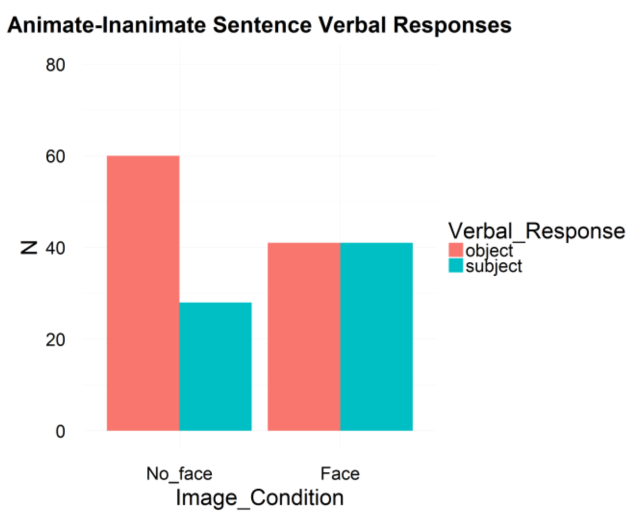

Figure 6: Child verbal response data by image condition for Animate-Inanimate sentences. N = number of verbal responses indicating pronoun referent.

Children: Post-hoc Analysis

In the model from Table 2, interactions between sentence type and image condition suggested that adding faces deceased strong preferences for the animate noun in sentences types such as IN_AN and AN_IN. We wanted to investigate whether or not adding faces removed differences between sentence types, and therefore overrode the effects of semantic animacy completely, or if visual animacy merely mitigated the effects of semantic animacy. To do so, the overall models from the eye-tracking and offline data were re-run, with AN_AN Face condition as the reference condition (in contrast to the initial models, where AN_AN NoFace was the reference). This allowed us to directly compare the sentence types with faces added to the AN_AN sentence type. The results of the main effects are reported in Table 6 below.

Table 6

Child eye-tracking and verbal response main effect results comparing sentence types to AN_AN Face intercept.

Sentence Type	**Model**	**ß**	**SE**	**t/z-score**	**p**
IN_IN	200-800 ms	-0.130	0.591	-0.220	0.827
	800-1400 ms	-0.616	0.981	-0.628	0.535
	1400-2000 ms	0.083	0.853	0.098	0.923
	verbal response	0.167	0.317	0.526	0.599
IN_AN	200-800 ms	-0.009	0.744	-0.012	0.990
	800-1400 ms	-1.26	1.026	-1.227	0.228
	1400-2000 ms	-0.586	1.165	-0.503	0.619
	verbal response	0.705	0.327	2.160	**0.031** *
AN_IN	200-800 ms	0.373	0.699	0.533	0.5967
	800-1400 ms	-0.126	1.136	-0.111	0.912
	1400-2000 ms	0.620	0.703	0.882	0.383
	verbal response	-0.199	0.3155	-0.630	0.529

$p < 0.05$ *. $p < 0.01$ **. $p < 0.0001$ ***

The results show that for eye-tracking data, adding faces removed any effects of semantic animacy on pronoun resolution. For the offline, verbal response data, the effect of adding faces was not strong enough to remove the differences between AN_AN and IN_AN sentences, as illustrated in the significant main effect. The effect of semantic animacy on AN_IN sentences in verbal response data disappeared when faces were added.

These results suggest that visual animacy modulated the processing of the sentences in the verbal response data, and that adding faces to images did not create equal processing to sentences. Adding faces to inanimate nouns did not override the effect of semantically inanimate nouns on pronoun resolution. Semantic animacy continued to affect pronoun resolution. Thus, it seems semantic animacy was a stronger cue than visual animacy in the strategic, offline responses.

In summary, both the child gaze data and response data illustrate that pronoun resolution was affected by both semantic and visual animacy. In general, participants were less likely to consider a noun as a referent if it was semantically inanimate. This semantic effect, however, was modulated by the introduction of visual animacy.

Adults: Eye-tracking Results

Overall analysis. The adult data were analyzed in the same fashion as the child data, except that covariates relating to age and Sentence Recall did not exist. Tables 7 and 8 present the eye-tracking results. As Table 7 shows, the reference condition, AN_AN NoFace, was not significantly difference from zero. That is, adult participants looked equally between the subject and object images in all three time frames. Within the sentence types in the NoFace condition, only AN_IN sentences (e.g., "…snake that the TV…") in the 3^{rd} time frame (1400-2000ms after pronoun onset) were significantly different from AN_AN sentences (e.g., "…snake that the bunny…"). In this time frame, the preference for an object antecedent (the snake) was stronger for the AN_IN sentences than the AN_AN sentences. Therefore participants considered "the TV" a less appropriate antecedent than "the bunny". Similar to the effect seen for IN_AN sentences for the child eye-tracking data, this illustrated an effect of semantic animacy on pronoun resolution in adults.

Similarly to the child data, there were significant interactions between sentence type and image condition, and a significant effect of trial order. Figure 7, below, shows that once again, participants displayed a decreased subject preference for IN_AN (yellow) sentences when faces were added, and a decreased object preference for AN_IN (green) sentences when faces were added in comparison to AN_AN sentences. Effects are illustrated in Figure 7 (the same figure with error bars included can be found in Appendix E).

Table 7

Adult eye-tracking results: Mixed-effects modelling of difference of logit looks to subject noun vs. object noun for three analysis time frames.

Predictor	**Time-frame**	**ß**	**SE**	**t**	**p**
Intercept	200-800 ms	0.591	0.760	0.778	0.443
	800-1400 ms	1.510	0.871	1.734	0.092
	1400-2000 ms	1.347	0.934	1.442	0.157
IN_IN	200-800 ms	-0.330	1.359	-0.243	0.810
	800-1400 ms	-1.150	1.053	-1.092	0.285
	1400-2000 ms	-0.332	1.029	-0.323	0.749
IN_AN	200-800 ms	1.639	1.291	1.269	0.218
	800-1400 ms	1.954	1.475	1.325	0.197
	1400-2000 ms	0.455	1.179	0.386	0.701
AN_IN	200-800 ms	-1.295	1.048	-1.236	0.226
	800-1400 ms	0.846	1.005	-0.842	0.404
	1400-2000 ms	-2.482	1.046	-2.373	**0.022** *
Face	200-800 ms	0.650	0.814	0.798	0.429
	800-1400 ms	0.305	0.789	0.386	0.701
	1400-2000 ms	-1.352	0.709	-1.907	0.063
Trial order	200-800 ms	-0.048	0.008	-5.899	**4.21e-09** ***
	800-1400 ms	-0.075	0.008	-9.144	**< 2e-16** ***
	1400-2000 ms	-0.024	0.008	-2.903	**0.004** **
IN_IN:Face	200-800 ms	-0.186	0.318	-0.586	0.558
	800-1400 ms	0.216	0.317	0.679	0.497
	1400-2000 ms	0.518	0.326	1.590	0.112
IN_AN:Face	200-800 ms	-1.505	0.293	-5.133	**2.97e-07** ***
	800-1400 ms	-1.438	2.963	-4.854	**1.26e-06** ***
	1400-2000 ms	0.167	0.308	0.543	0.587
AN_IN:Face	200-800 ms	1.201	0.299	4.013	**6.14e-05** ***
	800-1400 ms	0.684	0.302	2.267	**0.024** *
	1400-2000 ms	3.356	0.316	10.624	**< 2e-16** ***

$p < 0.05$ *. $p < 0.01$ **. $p < 0.0001$ ***

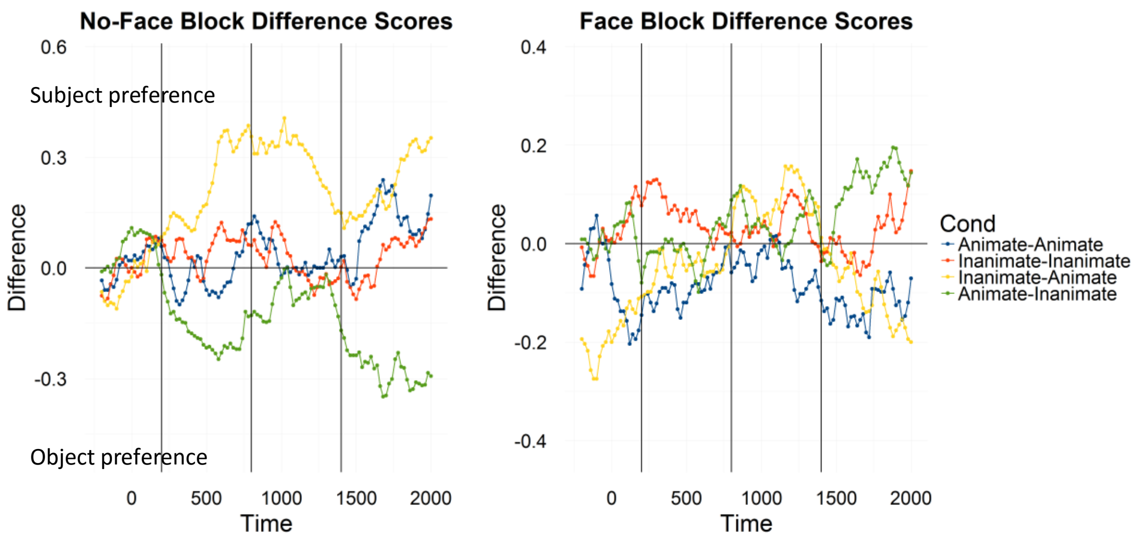

Figure 7. Adult eye-tracking data by sentence type and image condition. Difference scores = looks to subject interest areas subtract looks to object interest areas. Vertical lines indicate 200, 800 and 1400 ms respectively.

Table 8

Adult eye-tracking results: Mixed-effects modelling for individual sentence type.

Sentence Type	**Predictor**	**Time-frame**	**ß**	**SE**	**t**	**p**
Animate – animate	Intercept	200-800 ms	0.654	1.015	0.644	0.523
		800-1400 ms	4.123	1.156	3.566	**0.0009** ***
		1400-2000 ms	-1.044	1.175	-0.889	0.379
	Face	200-800 ms	-1.072	1.370	-0.782	0.440
		800-1400 ms	-0.189	1.136	-0.140	0.890
		1400-2000 ms	-1.679	1.332	-1.261	0.216
	Trial order	200-800 ms	-0.020	0.024	-0.811	0.418
		800-1400 ms	-0.181	0.026	-6.997	**8.17e-12*****
		1400-2000 ms	0.097	0.026	3.689	**0.0003*****
Inanimate – inanimate	Intercept	200-800 ms	-1.235	0.996	-1.240	0.225
		800-1400 ms	3.299	1.038	3.179	**0.003** **
		1400-2000 ms	4.439	1.052	4.221	**0.0001** ***
	Face	200-800 ms	1.577	1.242	1.269	0.222
		800-1400 ms	0.655	1.577	0.415	0.681

		1400-2000 ms	-0.500	1.616	-0.310	0.759
	Trial order	200-800 ms	0.029	0.015	1.977	**0.048** *
		800-1400 ms	-0.189	0.016	-12.093	**< 2e-16** ***
		1400-2000 ms	-0.220	0.015	-14.577	**< 2e-16** ***
Inanimate – animate	Intercept	200-800 ms	2.679	1.0315	2.598	**0.013** *
		800-1400 ms	1.545	0.831	1.859	0.068
		1400-2000 ms	2.090	0.814	2.567	**0.032** *
	Face	200-800 ms	-1.183	1.490	-0.794	0.433
		800-1400 ms	-1.368	1.532	-0.893	0.380
		1400-2000 ms	-1.543	1.361	-1.134	0.268
	Trial order	200-800 ms	-0.097	0.017	-5.625	**2.12e-08** ***
		800-1400 ms	0.004	0.017	0.242	0.809
		1400-2000 ms	-0.026	0.018	-1.465	0.144
Animate – inanimate	Intercept	200-800 ms	-0.003	0.849	-0.004	0.997
		800-1400 ms	-0.637	0.737	-0.865	0.391
		1400-2000 ms	-1.205	1.015	-1.187	0.242
	Face	200-800 ms	1.382	1.501	0.921	0.363
		800-1400 ms	1.728	1.308	1.321	0.195
		1400-2000 ms	1.560	1.215	1.317	0.197
	Trial order	200-800 ms	-0.044	0.018	-2.447	**0.015** *
		800-1400 ms	-0.004	0.017	-0.207	0.836
		1400-2000 ms	-0.004	0.019	-0.205	0.838

$p < 0.05$ *. $p < 0.01$ **. $p < 0.0001$ ***

Individual sentence-type analyses. In contrast to the child data, no sentence types showed a statistically significant effect of visual animacy, as seen in Table 8 (above). Trial order was significant in seven of the twelve models.

Post-hoc analysis. The absence of visual animacy main effects was surprising, when visually analyzing the data in Figure 7. When visually comparing the difference in effects of adding faces to the IN_AN sentence type to that of the AN_IN sentence type, it was clear that adding faces affected each sentence type in contrasting directions. To investigate this further, we ran another model containing the entire data set. It was identical to the original model in Table

7, except with IN_AN, NoFace set as the reference condition instead of AN_AN NoFace. This allowed us to look at the interaction between IN_AN and AN_IN sentences and visual animacy. The results are summarized in Table 9.

Table 9

Post-hoc adult eye-tracking results: Mixed-effects modelling of difference of logit looks to subject noun vs. object noun for three analysis time frames with IN_AN NoFace sentence type as intercept.

Predictor	**Time-frame**	**ß**	**SE**	**t**	**p**
Intercept	200-800 ms	2.230	1.223	1.823	0.084
	800-1400 ms	3.464	1.104	3.138	**0.003** **
	1400-2000 ms	1.802	7.358	2.449	**0.0193** *
AN_AN	200-800 ms	-1.639	1.291	-1.269	0.218
	800-1400 ms	-1.954	1.457	-1.325	0.197
	1400-2000 ms	-0.455	1.179	-0.386	0.701
IN_IN	200-800 ms	-1.969	1.620	-1.216	0.234
	800-1400 ms	-3.104	1.512	-2.053	**0.050** *
	1400-2000 ms	-0.788	1.177	-0.669	0.507
AN_IN	200-800 ms	-2.934	1.548	-1.895	0.0704
	800-1400 ms	-2.800	1.244	-2.252	**0.030***
	1400-2000 ms	-2.937	0.974	-3.014	**0.004** **
Face	200-800 ms	0.856	0.813	-1.053	0.298
	800-1400 ms	-1.133	0.791	-1.434	0.158
	1400-2000 ms	-1.185	0.702	-1.688	0.099
Trial order	200-800 ms	-0.048	0.008	-5.899	**4.21e-09** ***
	800-1400 ms	-0.075	0.008	-9.144	**< 2e-16** ***
	1400-2000 ms	-0.024	0.008	-2.903	**0.004** **
AN_AN:Face	200-800 ms	1.505	0.293	5.133	**6.81e-07*** **
	800-1400 ms	1.438	0.293	4.854	**1.26e-06** ***
	1400-2000 ms	-0.167	3.076	-0.543	0.587
IN_IN:Face	200-800 ms	1.319	0.293	4.503	**6.81e-06** ***
	800-1400 ms	1.654	0.297	5.561	**2.79e-08** ***
	1400-2000 ms	0.351	0.302	1.162	0.245

AN_IN:Face	200-800 ms	2.706	0.277	9.778	**< 2e-16** ***
	800-1400 ms	2.123	0.280	7.574	**4.11e-14** ***
	1400-2000 ms	3.189	0.289	11.052	**< 2e-16** ***

p < 0.05 *. p< 0.01 **. p <0.0001 ***

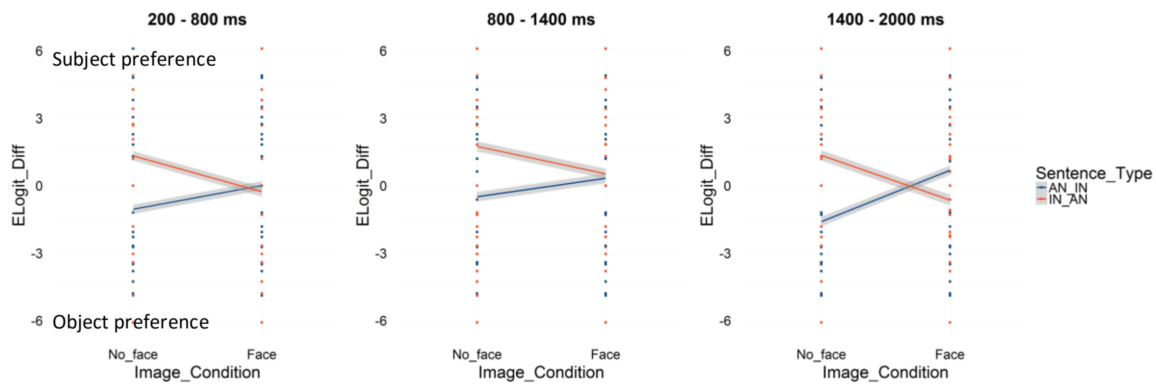

Figure 8. Sentence type by image condition interaction for inanimate-animate and animate-inanimate sentences. ELogit_Diff = looks to subject interest areas subtract looks to object interest areas.

As one can see in Table 9 under "AN_IN:Face", the interactions between the IN_AN and AN_IN sentences and visual animacy were significant in all three time frames. Adding faces consistently increased looks to the object in IN_AN sentences, where the object was inanimate, and increased looks to the subject in AN_IN sentences, where the subject was inanimate. Thus, adding eyes to images increased the amount they were considered by adults for pronoun antecedent. This interaction is illustrated in Figure 8 (above). Other main effects in Table 9 underline the differences in processing between IN_AN NoFace with the other sentence types, and other significant interactions underline the effect of visual animacy on the IN_AN Face condition. These effects were not considered central to the question at hand and so will not be further discussed here.

Adults: Offline Verbal Response Results

Overall analysis. The verbal responses to the comprehension questions for adults were also analysed in the same manner as the child data. We omitted the covariate of trial order from the final model due to non-significant effects on model fit. In contrast to the child verbal responses, but in line with the adult eye-tracking results, the adult participants showed no significant preference for object or subject antecedents in the reference condition, AN_AN NoFace. Tables 10 and 11 report the analysis results.

Table 10

Adult verbal response results: Mixed-effects modelling of verbal response to comprehension questions indicating subject noun vs. object noun interpretation of ambiguous pronoun.

Predictor	**ß**	**SE**	**z-value**	**p**
Intercept	0.010	0.359	0.028	0.978
IN_IN	0.366	0.442	0.083	0.409
IN_AN	2.267	0.558	4.066	**4.77e-05** ***
AN_IN	-0.914	0.447	-2.043	**0.041** *
Face	-0.109	0.448	-0.242	0.808
IN_IN:Face	0.376	0.622	0.604	0.546
IN_AN:Face	-1.199	0.713	-1.682	0.093
AN_IN:Face	1.196	0.610	1.959	0.050

$p < 0.05$ *. $p< 0.01$ **. $p <0.0001$ ***

Similar to the child verbal response data, in the model comparing all sentence types against the AN_AN sentences (e.g., "…snake that the bunny…") in the No Face condition both IN_AN (e.g., "…couch that the bunny…") and AN_IN (e.g., "…snake that the TV…") sentences showed significant differences, illustrating that there is an effect of semantic animacy on pronoun resolution when measured offline through a verbal response (Table 10). As illustrated in Figure 9, this trend is directionally similar to that observed for the children. For IN_AN sentences, there was an increase in participants indicating the animate subject (the bunny) as the antecedent. For AN_IN sentences, there was an increase in participants indicating that the animate object (the snake) was the antecedent. In contrast to the children, this difference disappeared when faces were added to the inanimate objects. That is, pronoun resolution was statistically the same as AN_AN sentences when all nouns were visually animate.

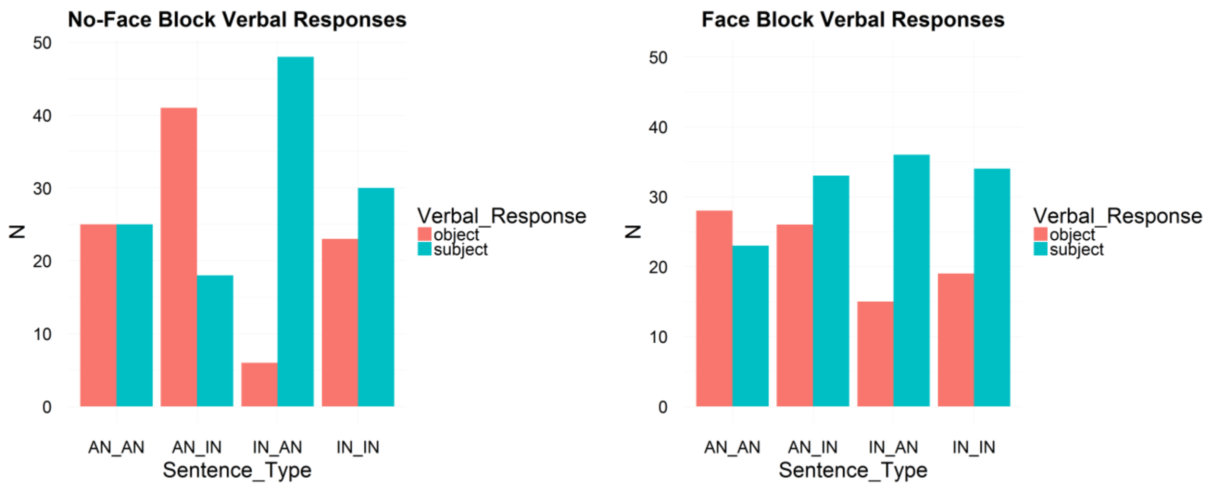

Figure 9: Adult verbal response data by sentence type and image condition. N = number of verbal responses indicating pronoun referent.

Table 11

Adult verbal response results: Mixed-effects modelling for individual sentence type.

Sentence Type	**Predictor**	**ß**	**SE**	**z-value**	**p**
Animate - animate	Intercept	-0.073	0.387	-0.188	0.851
	Face	-0.079	0.481	-0.164	0.870
Inanimate - inanimate	Intercept	0.266	0.277	0.959	0.338
	Face	0.316	0.399	0.793	0.428
Inanimate - animate	Intercept	2.103	0.469	4.487	**7.22e-06** ***
	Face	-1.217	0.542	-2.245	**0.025** *
Animate - inanimate	Intercept	-0.866	0.354	-2.45	**0.014** *
	Face	1.129	0.433	2.61	**0.009** **

$p < 0.05$ *. $p< 0.01$ **. $p <0.0001$ ***

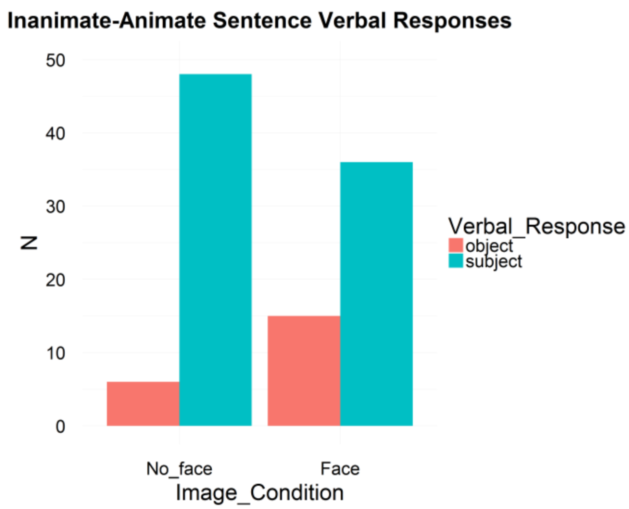

Figure 10: Adult verbal response data by image condition for Inanimate-Animate sentences. N = number of verbal responses indicating pronoun referent.

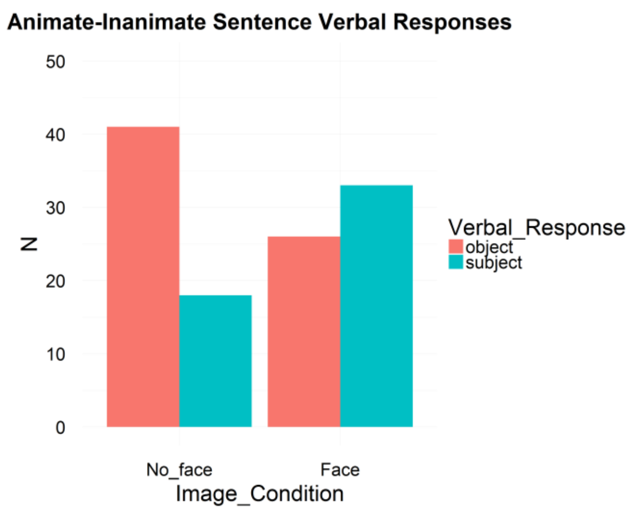

Figure 11: Adult verbal response data by image condition for Animate-Inanimate sentences. N = number of verbal responses indicating pronoun referent.

Individual sentence-type analyses. When looking at individual sentence types, one can see that in fact, visual animacy did significantly affect pronoun resolution for both sentence types IN_AN and AN_IN (Table 11, above). The effects are illustrated in Figures 10 and 11 (above). Similarly to the children, the subject preference in IN_AN sentences (e.g., "…couch that the bunny…") (Figure 10) decreases in strength when facial features are added to the object (the couch). For AN_IN sentences (e.g., "…snake that the TV…") (Figure 11), there is an object preference (the snake) in the NoFace condition, and equal preference in the Face condition.

Adult Post-Hoc Analysis

Similarly to the child data, we investigated the extent to which visual animacy modulated or outweighed effects of semantic animacy. To investigate if adding faces removed differences between sentence types for adults, and therefore overrode the effects of semantic animacy, the overall models from the eye-tracking and offline data were re-run, with AN_AN Face condition as the reference, similarly to the child data. This allowed us to directly compare the sentence types with faces added to the AN_AN sentence type. The results of the main effects are reported in Table 12 below.

The results mirror those for the child data. For eye-tracking data, adding faces removed any effects of semantic animacy. For the offline, verbal response data, adding faces was not sufficient to erase the differences between AN_AN and IN_AN sentences, yet was enough to remove the effect of semantic animacy on AN_IN sentences in verbal response data, as well as all sentence types in the eye-tracking data. Thus, semantic animacy appears to be a stronger cue than visual animacy for both adult and child participants only when they use offline strategies to resolve ambiguous pronouns.

Table 12

Adult eye-tracking and verbal response main effect results comparing sentence types to AN_AN Face intercept.

Sentence Type	**Model**	**ß**	**SE**	**t/z-score**	**p**
IN_IN	200-800 ms	-0.517	1.359	-0.380	0.706

	800-1400 ms	-0.935	1.056	-0.885	0.385
	1400-2000 ms	0.185	1.030	0.180	0.858
	verbal response	0.742	0.454	1.635	0.102
IN_AN	200-800 ms	0.133	1.290	0.103	0.919
	800-1400 ms	0.515	1.474	0.349	0.730
	1400-2000 ms	0.622	1.180	0.527	0.601
	verbal response	1.070	0.452	2.364	**0.018** *
AN_IN	200-800 ms	-0.094	1.045	-0.090	0.929
	800-1400 ms	-0.162	1.002	-0.162	0.872
	1400-2000 ms	0.875	1.043	0.838	0.40639
	verbal response	0.283	0.428	0.661	0.509

$p < 0.05$ *. $p < 0.01$ **. $p < 0.0001$ ***

In summary, the adult eye-tracking data show that semantic animacy did affect pronoun resolution. Although, there were no main effects of visual animacy, the significant interaction for IN_AN and AN_IN showed that visual animacy increased looks to inanimate nouns. A visual animacy effect is also apparent in the offline, verbal response data. The effects are similar to those seen with child participants. Again, nouns are less likely to be considered as a referent if they are semantically inanimate. This semantic effect is modulated by the introduction of visual animacy, so that visually animate nouns are more likely to be considered pronoun antecedents than the visually inanimate counterparts.

Overall Summary

The results overall showed patterns of preference for animate nouns over inanimate nouns as the antecedent for ambiguous pronouns. This was true for the child eye-tracking data in timeframe 1 (200-800 ms) for IN_AN sentence type. It was also true for the adult eye-tracking data in timeframe 3 (1400-2000 ms) for the AN_IN sentence type. Both children and adults displayed this animacy preference in their verbal responses for both IN_AN and AN_IN sentence types.

For visual animacy, the results overall displayed a preference for inanimate nouns with faces over inanimate nouns without faces. Child eye-tracking data revealed a main effect of visual animacy for AN_IN sentences, and interactions between sentence types and image conditions. Adult eye-tracking data found interactions between sentence types and image conditions both when AN_AN NoFace was reference, and when IN_AN NoFace was set as reference. Both children and adults displayed visual animacy main effects in their verbal responses for both IN_AN and AN_IN sentence types. These results displayed that adding faces increased preference for antecedent.

Discussion

The purpose of this study was to investigate how children would use semantic animacy and visual cues of animacy during language comprehension in comparison to adults. We hypothesized that children would use both semantic and visual cues of animacy to guide sentence interpretation similarly to adults when presented with a developmentally appropriate syntactic form. In order to test this, we used visual world eye tracking and stimulus items that took the form of sentences containing right branching, presentational-style object relative clauses, with animate and inanimate entities as grammatical subjects and objects of the sentences. All stimulus items violated some level of real-world pragmatic expectation in this study, as animals acted in anthropomorphic ways. Still, inanimate actors are semantically more unacceptable than animal actors. The hypothesis at hand was, if the images of inanimate nouns such as televisions and couches were given mouths and eyes, it should seem more acceptable for them to take an agentive role in the sentence. This acceptability would be demonstrated by an increase in assigning these inanimate nouns as referent to the ambiguous yet agentive pronoun, *he*. Thus, for both children and adults, we expected to see an effect of visual animacy on the proportion of looks to the subject noun vs. object noun during the presentation of the pronoun in the test sentence, as the visual context should have indicated the agentivity of the inanimate nouns. With respect to the hypotheses under investigation, the most important findings in this study were effects of semantic animacy and visual animacy for both children and adults in eye-tracking and offline data. We will address these findings and their implications to our understanding of how semantic and visual information is used in adult and child populations each in turn. We will then discuss additional findings, limitations to this study, and areas of future research.

Primary Findings

Table 1

Sentence Manipulations

	Head Noun (Object) Animacy	**Second Noun (Subject) Animacy**	**Sentence**
Sentence 1:	Inanimate	Inanimate	There is the couch that the TV phoned at school.
		Animate	There is the couch that the *bunny* phoned at school.
	Animate	Inanimate	There is the *snake* that the TV phoned at school.
		Animate	There is the *snake* that the *bunny* phoned at school.
Sentence 2:			He was excited to go to the movie that night.

Semantic animacy. The question for the first analysis was: would semantic animacy affect pronoun resolution? It has already been established that semantic animacy affects processing within comprehension of the relative clause itself (e.g., Brandt, et al., 2009; Clifton et al., 2003; Lowder & Gordon, 2014; Trueswell et al., 1994). However, it has not been established how semantic animacy affects the resolution of an animate, ambiguous pronoun, such as *he*. Nor have the effects of using semantically anomalous nouns been established. Nevertheless, semantic information such as gender and verb transitivity has been shown to affect pronoun resolution in adults and children (Arnold, Brown-Schmidt & Trueswell, 2007; Arnold et al., 2000; Pyykkönen et al., 2010; Rose, 2005). Like gender, animacy can be a distinguishing feature, as inanimate nouns should be referred to as *it,* and animate nouns should be referred to as *he/she*. Verb transitivity is one factor that increases semantic prominence. Factors that increase semantic prominence increase the saliency of certain nouns within sentences, which increases the likelihood that they will be referred to by pronouns in proceeding sentences. Like verb transitivity, animacy also can affect semantic prominence, as animate nouns are more likely to be agents, while inanimate nouns are more likely to be patients (Trueswell et al., 1994). These two factors suggest that animacy should be used during pronoun resolution by both adults and children.

In the current study, the expectation was that animate nouns would be more likely to be considered as antecedents than inanimate nouns. This was found for both the children and adults, but in different ways. For the child participants, this was established in both the eye-tracking data on the inanimate-animate sentences in the first time frame (200-800ms), and in the offline antecedent selection data for the inanimate-animate and animate-inanimate sentences. When the object of the initial sentence was inanimate, participants spent more time looking at the animate subject in reference to the pronoun, in comparison to when both nouns were animate (see Table 1, repeated above, for sentence examples). And, in the case of offline data, the participants were more likely to verbally indicate that the subject was the pronoun antecedent. When the subject of the initial sentence was inanimate, child participants were more likely to verbally indicate that the object was the pronoun antecedent. For the adult participants, the effect of semantic animacy was established in the eye-tracking data on the animate-inanimate sentences in the third time frame (1400-2000 ms), as well as in the offline data on the inanimate-animate and animate-inanimate sentences. The directions of effects followed the same patterns as the child data above. In sum, the effects of semantic animacy were seen in the children early in processing only when animacy matched subjecthood. For the adults, the effects were seen later in processing only when the animacy matched first-mention.

Inanimate-inanimate sentences never differed significantly from the animate-animate sentences, suggesting that a subject preference is a “default” approach to pronoun resolution until there is a contrast in animacy between the two nouns. Only then do we see that in general, pronoun reference is directed towards the animate noun.

The different patterns of semantic animacy effects between adults and children suggest underlying differences in processing between the two age groups. Currently, there is evidence that adults can use order-of-mention independently of subjecthood as early as 690-900 ms post-pronoun onset (favouring first mentioned as antecedent) (Järvikivi et al., 2005). Thus for adults, animacy may not be as reliable a cue in pronoun resolution as subjecthood and order-of-mention, therefore is only used later in processing as supportive information. Alternatively, the effect in the third time frame may reflect a conscious, strategic decision that the adult participants were making as they anticipated the comprehension question.

In the case of the child data, the animacy effect matching subjecthood (i.e., the effect found on IN_AN sentences) occurred in the first time frame (200 – 800 ms). This effect suggests a very early use of animacy in children more similar to their use of verb transitivity (significant as early as 200-720 ms in Pyykkönen et al., 2010) than their use of gender (significant as early as 600 – 800 ms in Arnold et al., 2007) or first-mention (significant after 1000 ms in Song & Fisher, 2007 and Pyykkönen et al. 2010) during pronoun resolution. Thus, animacy may be used, at least in the case of implausible sentences, as another cue to semantic prominence in children. It is unlikely that animacy is supporting a preference for subject antecedent in the children in the manner we suggest order-of-mention is used for the adults because all research on child pronoun resolution suggests that any possible use of subjecthood/order-of-mention is later occurring in children, 1240 – 1760 ms after pronoun onset (Pyykkönen et al., 2010).

To sum, it appears that adults rely primarily on order-of-mention/subjecthood during pronoun resolution, and supplement with semantic animacy during later stages of processing. Children, who do not yet use order-of-mention/subjecthood in adult-like ways, may rely on semantic animacy as an early cue to semantic prominence and agentivity. Though the exact mechanisms for how adult and child participants use semantic animacy require further exploration, it is nonetheless true that effects were found, both in the eye-tracking data and more prominently in the offline verbal response data. All significant effects showed animate nouns were preferred as antecedents over inanimate nouns.

Visual animacy. The effect of semantics established, we then looked at the individual sentence types to compare the images without faces to the images with faces added to determine whether the manipulation of visual animacy modulated pronoun interpretation. For the child eye-tracking data, an effect of visual animacy was seen for the animate-inanimate sentence type in both the second and third time frames (800-1400 ms and 1400-2000 ms) as indicated in Table 3 of the results section. Without faces, children looked significantly more to the object image (in comparison to chance). When faces were added this object preference was eliminated. That is, the children looked equally to both the object and subject images. There was no significant main effect of visual animacy in the adult eye-tracking data (see “Face” conditions in Table 8), but interactions between sentence types and image condition indicated that visual animacy continued to modulate pronoun resolution, as indicated by AN_IN:Face in Table 9. Additionally, both child and adult groups showed a significant main effect of visual animacy for both inanimate-animate and animate-inanimate sentences in the offline verbal response data. In all cases, the direction of effects suggests that adding faces to images increased the chance that inanimate nouns would be considered the referent of the pronoun when an explicit response was required. These results demonstrate that online, the children relied more heavily on the visual information than the adults, and offline, adults and children eventually arrived at similar conclusions regarding pronoun resolution.

An interesting pattern to note is that in general the offline data, verbal responses to comprehension questions, were more sensitive to the effects of both semantic and visual animacy. For both age groups, the eye tracking data revealed significant effects of semantic animacy for one sentence type only and within one time frame. Moreover, in the adult eye tracking data, there were no significant main effects of visual animacy at all. Yet for both groups, semantic and visual animacy effects were much more robust in the offline comprehension data.

It may be that animacy (both semantic and visual) is used only marginally in adults, as they are able to rely more strongly on other cues to agentivity during pronoun resolution (Järvikivi, Van Gompel & Hyönä, 2016, in press; Schumacher, Roberts & Järvikivi, in press). Children on the other hand, perhaps due both to overall less linguistic experience, and proportionally more experience with fictional situations, use animacy cues online to determine agentivity. Both groups then rely more heavily on animacy cues in strategic ways offline, when presented with a forced-choice decision. Indeed, during data collection, it was noted that many participants (both child and adult) were initially confused by the comprehension questions, and uncomfortable in making a decision on antecedent. Despite this discomfort, participants in both age groups made decisions in a way that appears to be systematic. This is an example of the difference between online and offline data. Eye-tracking data allow us to investigate how the manipulations affect moment to moment interpretations of language, from the moment the pronoun is encountered, whereas verbal response data allows us to investigate how the same manipulations affect final, explicit interpretations of the language. The fact that children showed main effects of visual animacy in their eye-gaze data when adults did not speaks to the automatic reliance of young children on non-linguistic information over the linguistic information, whereas adults use it in conjunction with subjecthood and order-of-mention. This falls in line with the findings of Armstrong, Andreu, Esteve-Gibert and Prieto (2016) who found that if visual context was provided before the presentation of the linguistic information, children based their interpretations on the visual information, and had more difficulty integrating the linguistic information into their final interpretations. Armstrong et al. (2016) argued that children did not have difficulties using the visual context, but in revising initial hypotheses. In this study, because children saw the cartoon pictures before onset of the language, it may have increased their ability to use the information during their comprehension. Additionally, adults appeared to automatically rely more heavily on the linguistic information. This by-group difference may be due in part to the difference in distributions of linguistic and non-linguistic experiences between the two groups. One may consider that the difference in experiences lead adults to inhibit the visual information, as in their overall day-to-day experiences, having eyes is not a strongly relevant cue to pronoun resolution. Because children are often exposed to cartoons with inanimate agents who have faces, visual animacy may be more relevant information during pronoun resolution in comparison to adults. Alternatively, the visual information may not be more relevant to the children; they may just be less able to inhibit it during pronoun resolution (Järvikivi & Pyykkönen-Klauck, in submission).

This study differs from previous research on how visual context can affect language comprehension in children in two major ways. First, many previous studies investigated whether visual context affected syntactic comprehension in children (Hurewitz et al. 2000; Snedeker & Trueswell, 2004; Trueswell et al. 1999; Weighall & Altmann, 2011; Zhang & Knoeferle, 2012). In contrast, this study investigated whether visual context affected use of semantic information in comprehension. Other studies that focused on semantic interpretations also found effects of visual context on processing (Huang and Snedeker, 2008; Van Rij et al., 2016). While children's use of visual cues during syntactic processing remains under debate, it appears much clearer that semantic processing is affected by visual context. Second, previous studies used linguistic items that had only one correct interpretation to observe possible effects of visual information, whereas this study used ambiguous pronoun resolution, where structurally, both subject and object reference may have been possible. Using ambiguous pronouns allows us to observe the effects of various cues on language processing without requiring participants to formulate and revise hypotheses on sentence structure. Thus, finding effects aids in verifying that such cues are used at all times during processing, rather than in cases where syntactic cues are not enough on their own.

Semantic information is learned initially through connecting words to our real-world experiences that are accessed through visual and other sensory processes (Gleitman, Cassidy, Nappa, Papafragou & Trueswell, 2005). It would be advantageous then, for semantic interpretations to be flexible to changes in non-linguistic context. Nieuwland and Van Berkum (2006) have suggested that our semantic representations of lexical items are not separate from pragmatic knowledge, how words are used in different situations. They used discourse context, text setting the stage for a fictional situation, to illustrate this. The current study supports and extends their suggestion, by showing that non-linguistic visual information affects how semantic information is used during comprehension. Perhaps the visual information constrains the pragmatic context to one of a fictional, cartoon scenario in the same way discourse context did so for the Nieuwland and Van Berkum (2006) study, where a fictional scenario changed what was considered semantically appropriate in language. The ability for "semantic appropriateness" to change with context suggests that it is not inherent to the language system, but tied closely to pragmatic and non-linguistic context. However, semantic animacy remained a strong cue within and of itself in the offline data, as seen by significant differences between the AN_AN and IN_AN sentences within the faces block. Thus, when faced with conscious decisions about ambiguous pronouns, both adults and children still use semantic animacy to some degree separately from visual animacy.

The current study shows that this semantic flexibility is characteristic of both adults and children. In fact, the effects appear to be more robust in children, as significant main effects of visual animacy were available for both eye-tracking and offline data, whereas for adults the results were statistically significant for only interactions and offline data. This may indicate a stronger, online learning element to the children's processing, as they may still be building semantic representations of lexical items. Alternatively, they may be adapting more quickly to the fictional context due to the types of situations children see in their day-to-day life. Children often watch cartoon television, are read story books, and incorporate fictional situations in their imaginary play where inanimate objects take animate roles. Therefore, it may be easier for them to switch pragmatic contexts immediately than adults, who are generally exposed to more realistic situations, even in literature and media. A third alternative, discussed above, is that adults integrate the animacy information with cues of subjecthood and order-of-mention online, whereas children cannot integrate such information online, and generally have not fully acquired early use of subjecthood/order-of-mention (Arnold et al., 2007; Song & Fisher, 2007; Pyykkönen et al., 2010).

Additional Findings

Throughout the eye-tracking data for both participant groups, there was a consistently significant effect of trial order on pronoun resolution. In this experiment, there were five types of models (one full model to establish effects of semantic animacy, and a separate model for each sentence type to find effects of visual animacy) run on three separate time frames for each participant group, summing up to 30 models for eye-tracking data. Of those 30 models, trial order was significant 23 times. However, the directionality of this effect varied between positive, increasing looks to subject, and negative, increasing looks to object. While at first this effect appeared random, closer inspection revealed a relationship with the directionality of the intercept, that is, the reference condition, illustrated in Tables 13 and 14. In all but two of the 23 cases of significant trial order, the directionality of trial order was opposite of the directionality of the intercept. The directionality of the intercept indicates the pronoun preference for the comparison sentence: either significantly more looks to subject, significantly more looks to object, or looks not significantly different from chance. That means that when there were significantly more looks to the subject image, over the course of the experiment looks to the object increased. When there were significantly more looks to the object image, over the course of the experiment looks to the subject increased. In the two circumstances where this pattern was not the case, the intercept was not significantly different from chance. This pattern suggests that over the course of the experiment participants became less sure in their pronoun resolution. Likely this is due to the unpredictability of the animacy of the nouns throughout the experiment. Participants adapted to the task and became more cautious and deliberate in their pronoun resolution.

Table 13

T-score directionality of intercepts/trial order for child eye-tracking data.

"+" = significantly more looks to subject; "-" = significantly more looks to object; "0" = not statistically significant.

Time Frame (ms)	**Full Model (Table 2)**	**AN_AN (Table 3)**	**IN_IN (Table 4)**	**IN_AN (Table 5)**	**AN_IN (Table 6)**
200-800	-/+	-/+	0/-	0/+	-/+
800-1400	0/-	+/-	+/-	+/-	-/+
1400-2000	0/0	0/+	0/0	0/+	0/-

Table 14

T-score directionality of intercepts/trial order for adult eye-tracking data.

"+" = significantly more looks to subject; "-" = significantly more looks to object; "0" = not statistically significant.

Time Frame (ms)	**Full Model (Table 12)**	**AN_AN (Table 13)**	**IN_IN (Table 14)**	**IN_AN (Table 15)**	**AN_IN (Table 16)**
200-800	0/-	0/0	0/+	+/-	0/-
800-1400	0/-	+/-	+/-	0/0	0/0
1400-2000	0/-	0/+	+/-	+/0	0/0

Age and Recalling Sentences scores did not significantly aid prediction of pronoun resolution in children. This differs from Weighall and Altmann's (2011) study, which found differences in use of visual information for children aged six to eight with high vs. low verbal memory spans, suggesting that factors that may affect cognitive capacity could play a role in children's abilities to use visual information during language processing. The authors reviewed information that suggests that higher memory spans allow children to have more cognitive resources to incorporate the non-linguistic information. It is possible that in this case, combined use of semantic information and contextual information is mastered before the age and ability of the participants in our study. Considering the more recent discovery by Armstrong et al. (2016)

that children have difficulty revising hypotheses regardless of the type of information, increased cognitive capacity may have allowed the children in Weighall and Altmann's study the ability to complete such revisions when needed. In this particular study, participants were not required to create and revise interpretation hypotheses to resolve the pronouns. Thus, variations in cognitive capacity, as measured by age or Recalling Sentence scores would be less relevant to the task. A third possibility to be considered is a lack of variability in the covariates sufficient to expose individual difference in cognitive capacity. However, this seems unlikely, as ages ranged from 48 to 63 months old, and Recalling Sentences scores ranged from 7 to 18.

Finally, a secondary benefit of using relative clauses to investigate semantic and visual animacy effects on pronoun resolution is increased knowledge of the interplay between subjecthood and order-of-mention in spoken English comprehension through analysis of AN_AN sentences. In children, an early preference (200-800 ms) for first-mention was seen in eye-tracking data. This preference was not maintained, and in offline, verbal response data, they displayed a subject preference. These effects should be corroborated and further investigated, as current pronoun resolution research has found only very late subjecthood/order-of-mention effects (Song & Fisher, 2007 and Pyykkönen et al., 2010). Indeed, these studies used sentences where first-mention and subjecthood matched, and thus targeted the same noun as antecedent, whereas in this study, they did not. The adults, on the other hand, showed no preference in AN_AN sentences. This suggests that for English (like in Finnish, Järvikivi et al., 2005) order-of-mention and subjecthood have separate effects on pronoun resolution in adult processing.

Limitations

There are several limitations to this study. First, both the child and adult participant groups were fairly homogenous in terms of socioeconomic status (SES). All but four of the children's mothers had a college or university degree or higher. The families participating were able to afford for their children to attend a daycare/preschool, or had the means to bring their children into the university lab, which also indicates a generally higher SES of the group. SES may affect the type and frequency of language children are exposed to, both through interactions with people and media such as books and television (Hoff, 2003). All adult participants were attending university. This is also a fairly homogenous group in terms of age, SES, and language background. Homogeneity of participant groups makes generalization of results to the larger population more difficult overall. As with all research, findings should ideally be replicated with experimental groups that reflect the larger population.

Second, it is also necessary to acknowledge that there is no way to distinguish the level of confidence in the answers participants gave to the comprehension questions. As mentioned briefly above, it was noted that some participants initially had a hard time understanding how to answer the comprehension questions. In some instances, participants may have been able to come to a final interpretation of the pronoun they were confident with. In other instances, participants may have felt that they never truly decided which noun was the antecedent. It is possible that some or all participants tried to guess what answer was expected from them rather than truly believing that a particular noun was the antecedent. However, half of the comprehension questions were filler questions with right and wrong answers. This high proportion of filler questions should have mitigated the chances of participants randomly guessing, and allowed the exclusion of participants who guessed consistently. And whether final interpretations were confident or not, it is clear the experimental manipulations affected the outcomes of their responses.

Future Research

Future research may replicate and extend the findings of this study. Inclusion of a measure of exposure to fictional cartoon situations would be of interest for both children and adult participants. Children with limited exposure to cartoon television and literature may show differences in processing to their highly exposed counterparts. There may also be differences in processing between university students and adults such as parents, caregivers and educators of small children, who are exposed to much the same situations as the children they care for. Such findings would help distinguish if differences between adults and children are due to the relevance of the visual information in to the individuals' lives, or if the differences are due to children's inability to inhibit the irrelevant visual information. EEG studies for both adults and children may also expose how visual information affects semantic interpretations. For example, would size of the N400, a semantic event-related potential, be affected by visual context, similarly to how it was affected by discourse by Nieuwland and Van Berkum (2006)? How does this vary between populations?

Importantly, it is also pertinent to investigate the differences in processing between typically developing children and children with specific language impairment (SLI). One of many explanations of SLI suggests impairment is due to reduced processing capacity which negatively affects the acquisition and use of language forms (Bishop, 1994; Im-Bolter, Johnson & Pascual-Leone, 2006; Leonard et al., 2000). The limited processing capacity account suggests that for any task, as cognitive demands increase, the child's limited capacity cannot maintain optimal performance, and language performance is affected. What does this mean for integration of semantic and visual cues during language processing for the SLI population? On one hand, providing a visual cue may reduce the cognitive load of processing auditory language, making performance more similar to typically developing children. On the other hand, the process of integrating information from different sources may increase the cognitive demands of the task, reducing the child's abilities to process the language. Weighall & Altmann (2011) found that typically developing children with higher memory spans gained more benefit from visual context when processing complex sentences than children with lower memory spans. It is well documented that children with SLI often have reduced working memory abilities relative to peers (Lum, Conti-Ramsden, Page & Ullman, 2012; Montgomery, 2003). If SLI is due to a demands and capacity trade off, children with SLI may not benefit from visual cues. Investigation of how children with SLI use visual information in conjunction with semantic information would both shed light on the mechanisms behind language impairments, and affect how speech-language pathologists and educators use visual information to support language learning in this population.

Lastly, a secondary finding to this study, the early online first-mention preference in children, the offline subject preference in children, and the overall lack of preference in adults for typical, animate-animate sentences highlights need for continued research on pronoun resolution for subjecthood and order-of-mention in both children and adults. Relative clauses offer a unique opportunity in English to separate the two constraints, which otherwise generally go hand-in-hand in English. Further research where animacy is not manipulated in surrounding test sentences, and subject relative clauses are included, may aid in exploring how subjecthood and order-of-mention interact in adult and child pronoun resolution.

Conclusion

This study aimed to shed light on how adults and children use visual and semantic animacy when resolving pronouns following right-branching object relative clauses. We initially hypothesized that both children and adults' pronoun resolution would be affected by semantic animacy, showing a preference for animate antecedents. We also hypothesized that both children and adults would be affected by visual animacy, showing that adding eyes and mouths to images of objects would increase preference for inanimate nouns as antecedents. Additionally, we expected to see these effects both in online, eye-tracking data, as well as offline, verbal response data.

In fact, we did find effects of semantic animacy on pronoun resolution in both children and adults. These effects occurred for sentences with either subject or object animacy variations, over different time-frames in the eye-tracking data. This suggests that children and adults use the same information through different processes. We suggest that adults rely more heavily on subjecthood/order-of-mention constraints, and use semantic animacy as supporting information to confirm their interpretations. We suggest that children, who do not appear to be able to use such constraints as adults do during pronoun resolution, use semantic animacy immediately as a cue for semantic prominence. We also found effects of semantic animacy on verbal response data for both adults and children. All effects suggest that animate nouns are preferred over inanimate nouns during pronoun resolution.

We found main effects of visual animacy on pronoun resolution in eye-tracking data for children only. Adult use of visual animacy was supported through sentence type and image condition interactions only. Both groups were more likely to consider inanimate nouns as antecedents if the corresponding image had facial features. It appears that for adults, the inclusion of facial features was not a strong enough constraint to override linguistic constraints on pronoun resolution to the extent of creating main effects. We suggest that this is due to the infrequency of anthropomorphized objects in most adults' everyday life in comparison to children, or may instead be due to the children's inability to integrate the visual information with other semantic information. However, both groups displayed effects of visual animacy in the verbal response data. Therefore, adults are capable of using visual information of animacy in a more explicit, higher-level task, such as answering questions.

The findings of this study underscore not only that children are capable of using visual information during language comprehension, but that there exist fine variations between adult and child processing. Semantic and visual animacy affected the final interpretations of pronouns in the same ways for adults and children. Yet online, the adults and children used this information in very different ways. These different uses of information may be due, in part, to difference in proportional exposure to types of situations across the two groups. Future research is necessary to confirm this hypothesis. Findings from this study support constraint-based theories of language development that include mechanisms for children to use both linguistic and non-linguistic cues immediately online, though in ways sometimes different from adults.

References

Armstrong, M. E., Andreu, L., Esteve-Gibert, N., & Prieto, P. (2016). Children's processing of morphosyntactic and prosodic cues in overriding context-based hypotheses: an eye tracking study. *Probus*, *28*(1), 57-90.

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22*(4), 527-565.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition, 76*, B13-B26.

Arnon, I. (2005). Relative clause acquisition in Hebrew: Towards a processing-oriented account. In *Proceedings of the 29th Boston University conference on language development* (Vol. 29, pp. 37-48). Somerville, MA: Cascadilla Press.

Bååth, R. (2010). ChildFreq: An Online Tool to Explore Word Frequencies in Child Language. *LUCS Minor*, 16.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457-474.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Bishop, D. V. (1994). Grammatical errors in specific language impairment: Competence or performance limitations? *Applied Psycholinguistics*, *15*(04), 507-550.

Brandt, S., Kidd, E., Lieven, E., & Tomasello, M. (2009). The discourse bases of relativization: An investigation of young German and English-speaking children's comprehension of relative clauses. *Cognitive Linguistics*, *20*(3), 539-570.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687.

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language*, *65*(2), 128-144.

Clikr. [online images]. Retrieved from http://www.clker.com/ January 2015 through February 2015.

Coco, M. I., & Keller, F. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *Quarterly Journal Of Experimental Psychology, 68*(1), 46-74. doi:10.1080/17470218.2014.936475

Corrêa, L. M. S. (1995). An alternative assessment of children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *24*(3), 183-203.

Clifton Jr, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*(3), 317-334.

Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language, 81*(4), 882-906.

Ferreira, F., Foucart, A., & Engelhardt, P. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language, 69*(3), 165-182.

Flaming Text (2015). ImageBot, http://www.flamingtext.com/imagebot/ .

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior, 22*(2), 203-218.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23-64.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368-1378.

Huang, Y. T., & Snedeker, J. (2008). Use of referential context in children's language processing. In *Proceedings of the 30th annual meeting of the Cognitive Science Society* (pp. 1380-1385).

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*(6), 597-626.

Im-Bolter, N., Johnson, J., & Pascual-Leone, J. (2006). Processing limitations in children with specific language impairment: The role of executive function. *Child Development*, *77*(6), 1822-1841.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446.

Järvikivi, J., Van Gompel, R. P. G., Hyönä, J. (2016, in press). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *Journal of Psycholinguistic Research*.

Järvikivi, J., Van Gompel, R. P., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution contrasting the first-mention and subject-preference accounts. *Psychological Science*, *16*(4), 260-264.

Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., & Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: tracking eye movements to visually presented anaphoric referents. *Language, Cognition and Neuroscience*, *29*(7), 877-892.

Järvikivi, J., & Pyykkönen-Klauck, P. (submitted). Changes in referent co-presence influence pronoun comprehension in children.

Kidd, E., & Bavin, E. L. (2002). English-Speaking Children's Comprehension of Relative Clauses: Evidence for General-Cognitive and Language-Specific Constraints on Development. *Journal of Psycholinguistic Research*, *31*(6), 599–617. doi:10.1023/A:1021265021141

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, *22*(6), 860–897. doi:10.1080/01690960601155284

Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery. *Journal of Child Language*, *38*(01), 222-234.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*(5), 580-602.

Leonard, L. B., Miller, C. A., Grela, B., Holland, A. L., Gerber, E., & Petucci, M. (2000). Production operations contribute to the grammatical morpheme limitations of children with specific language impairment. *Journal of Memory and Language*, *43*(2), 362-378.

Lowder, M. W., & Gordon, P. C. (2014). Effects of animacy and noun-phrase relatedness on the processing of complex sentences. *Memory & Cognition*,*42*(5), 794-805.

Lum, J. A., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*(9), 1138-1154.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*(1), 50-68.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372-380.

Montgomery, J. W. (2003). Working memory and comprehension in children with specific language impairment: what we know so far. *Journal of Communication Disorders*, *36*(3), 221-231.

Nappa, R., Wessel, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, *5*(4), 203-234.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When Peanuts Fall in Love: N400 Evidence for the Power of Discourse. *Journal of Cognitive Neuroscience, 18*(7), 1098–1111.

OIE Internet Concepts. Online Image Editor, http://www.online-image-editor.com/ .

Openclipart (2004). [online images]. Retrieved from https://openclipart.org/search/?query=cartoon March, 2014 through February 2015.

Porretta, V., Kyröläinen, A, Van Rij, J., & Järvikivi, J. (2016). *VWPre: Tools for Preprocessing Visual World Data*. R package version 0.7.0.

Pyykkönen, P., Matthews, D., & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes, 25*(1), 115129. doi:10.1080/01690960902944014

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413-425.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748. doi:10.1111/1469-7610.00770

Rose, R. L. (2005). The relative contribution of syntactic and semantic prominence to the salience of discourse entities. Unpublished PhD, Northwestern University.

Schumacher, P., Roberts, L., & Järvikivi, J. (in press). Agentivity drives real-time pronoun resolution: Evidence from German er and der. *Lingua*.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*(4), 569-588.

Shutterstock Inc. (2015). [online images]. Retrieved from http://www.shutterstock.com/ February 2015.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*(3), 238-299.

Song, H., & Fisher, C. (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua, 117*(11), 1959-1987. doi:10.1016/j.lingua.2006.11.011

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447-481.

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*(3), 227-267.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47*(1), 69-90.

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

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*(2), 89-134.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*(3), 285-318.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 528.

Van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. Holler, A., Goeb, C., & Suckow, K. (Eds.), *Experimental Perspectives on Anaphora Resolution. Information Structural Evidence in the Race for Salience*, 242-267

Vogels, J., Krahmer, E., & Maes, A. (2013). When a stone tries to climb up a slope: the interplay between lexical and perceptual animacy in referential choices. *Frontiers in Psychology, 4*.

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, *36*(05), 559-570.

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, *99*(2), 75-95.

Weighall, A. R., & Altmann, G. T., (2011). The role of working memory and contextual constraints in children's processing of relative clauses. *Journal of Child Language, 38*(3), 579-605.

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical evaluation of language fundamentals—Preschool, second edition (CELF Preschool-2)*. Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Zhang, L., & Knoeferle, P. (2012). Visual context effects on thematic role assignment in children versus adults: Evidence from eye tracking in German. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. *Austin TX: Cognitive Science Society* (pp. 2593-2598).

Appendix A

UNIVERSITY OF ALBERTA

DEPARTMENT OF COMMUNICATION SCIENCES & DISORDERS

2-70 Corbett Hall
Edmonton, Alberta, Canada T6G 2G4
Tel: 780.492.5990
Fax: 780.492.9333
www.csd.ualberta.ca

BACKGROUND QUESTIONNAIRE

Study Title: Cartoons and comprehension: The effect of visual context on children's sentence processing

Date of birth: Day: ___________ Month: _____________ Year: ________________

(1) What is your first and primary language?

(2) Do you hear or speak a language other than English on a regular basis?

Yes ☐ No ☐

a) If yes, please provide us more information about how often and for how long:

b) How well do you speak each language? Please check a white box on the scale:

English

Very little English ←	1	2	3	4	5	Very fluent in English →

Other: ___________

Very little ←	1	2	3	4	5	Very fluent →

(3) Do you have normal hearing? _______________________

(4) Do you have normal or corrected-to-normal vision? ______________________

Appendix B

UNIVERSITY OF ALBERTA

DEPARTMENT OF COMMUNICATION SCIENCES & DISORDERS

2-70 Corbett Hall
Edmonton, Alberta, Canada T6G 2G4
Tel: 780.492.5990
Fax: 780.492.9333
www.csd.ualberta.ca

PARENT QUESTIONNAIRE

Study Title: Cartoons and comprehension: The effect of visual context on children's sentence processing

Child's date of birth: Day: ___________ Month: _____________ Year: ________________

(1) What is your child's first and primary language? ____

(2) Does your child hear or speak a language other than English on a regular basis? Yes ☐ No ☐

a) If yes, please provide us more information about how often and for how long (e.g., attends French Immersion school, attends language class 2 hours/week, speaks with grandparents during weekend visits, etc.):

____

____

b) How well does your child speak each language for his/her age? Please check a white box on the scale:

English

Very little English ←	1	2	3	4	5	Very fluent in English →

Other: ___________

Very little ←	1	2	3	4	5	Very fluent →

(3) Do you have any concerns about your child's language development, or does your child have any learning difficulty or condition that would affect his or her ability to use language? Yes ☐ No ☐

If yes, please provide us with more information (optional):

____

____

(4) Does your child have normal hearing? _______________________
(5) Does your child have normal or corrected-to-normal vision? ______________________

(6) What is the child's mother's highest level of education?
☐ Some high school ☐ Some college / university ☐ Some graduate school / advanced degree (eg, MD)
☐ High school ☐ College / university ☐ Graduate school / advanced degree

Appendix C

Test Sentence Pairs – Four versions of sentence 1 and one version of sentence 2 (in bold)

1. This is the tree that the ball fed in the living room.
   This is the tree that the duck fed in the living room.
   This is the lion that the ball fed in the living room.
   This is the lion that the duck fed in the living room.
   He was on his best behaviour that day.

2. Here is the TV that the tree pinched in the kitchen.
   Here is the TV that the lion pinched in the kitchen.
   Here is the bunny that the tree pinched in the kitchen.
   Here is the bunny that the lion pinched in the kitchen.
   He was not acting very nice yesterday.

3. There is the couch that the TV phoned at school.
   There is the couch that the bunny phoned at school.
   There is the snake that the TV phoned at school.
   There is the snake that the bunny phoned at school.
   He was excited to go to the movie that night.

4. That is the flower that the couch cuddled in the treehouse.
   That is the flower that the snake cuddled in the treehouse.
   That is the pig that the couch cuddled in the treehouse.
   That is the pig that the snake cuddled in the treehouse.
   He was very tired but very happy.

5. This is the cake that the flower squashed in the bathroom.
   This is the cake that the pig squashed in the bathroom.
   This is the horse that the flower squashed in the bathroom.
   This is the horse that the pig squashed in the bathroom.
   He felt frustrated and didn't know what to do.

6. Here is the bus that the cake kissed at the lake.
   Here is the bus that the horse kissed at the lake.
   Here is the turtle that the cake kissed at the lake.
   Here is the turtle that the horse kissed at the lake.
   He ran into the living room right away.

7. There is the orange that the bus squeezed at the park.
   There is the orange that the turtle squeezed at the park.
   There is the elephant that the bus squeezed at the park.
   There is the elephant that the turtle squeezed at the park.
   He wasn't playing very fair at all.

8. That is the carrot that the orange kicked in the yard.
   That is the carrot that the elephant kicked in the yard.
   That is the frog that the orange kicked in the yard.
   That is the frog that the elephant kicked in the yard.
   He was really mad all day long.

9. This is the apple that the carrot banged at the farm.
   This is the apple that the frog banged at the farm.
   This is the monkey that the carrot banged at the farm.
   This is the monkey that the frog banged at the farm.
   He had been trying really hard to calm down.

10. Here is the truck that the apple hit in the store.
    Here is the truck that the monkey hit in the store.
    Here is the owl that the apple hit in the store.
    Here is the owl that the monkey hit in the store.
    He had been causing a lot of trouble.

11. There is the chair that the truck bumped near the forest.
    There is the chair that the owl bumped near the forest.
    There is the giraffe that the truck bumped near the forest.
    There is the giraffe that the owl bumped near the forest.
    He kept tripping all over the place.

12. That is the banana that the chair teased outside.
    That is the banana that the giraffe teased outside.
    That is the dog that the chair teased outside.
    That is the dog that the giraffe teased outside.
    He went inside afterwards to play.

13. This is the car that the banana found behind the house.
    This is the car that the dog found behind the house.
    This is the cow that the banana found behind the house.
    This is the cow that the dog found behind the house.

He had been wandering around all day.

14. Here is the shoe that the car loved by the mountain.

    Here is the shoe that the cow loved by the mountain.
    Here is the mouse that the car loved by the mountain.
    Here is the mouse that the cow loved by the mountain.
    He wanted to be best friends forever.

15. There is the pen that the shoe hated at the mall.
    There is the pen that the mouse hated at the mall.
    There is the bear that the shoe hated at the mall.
    There is the bear that the mouse hated at the mall.
    He had gotten very upset last week.

16. That is the cup that the pen ignored by the pond.
    That is the cup that the bear ignored by the pond.
    That is the cat that the pen ignored by the pond.
    That is the cat that the bear ignored by the pond.
    He was having a very long week.

17. This is the hat that the cup liked at the river.
    This is the hat that the cat liked at the river.
    This is the chicken that the cup liked at the river.
    This is the chicken that the cat liked at the river.
    He thought they would become good friends.

18. Here is the box that the hat heard down the road.
    Here is the box that the chicken heard down the road.
    Here is the wolf that the hat heard down the road.
    Here is the wolf that the chicken heard down the road.
    He asked what was going on.

19. There is the strawberry that the box lost at the fair.
    There is the strawberry that the wolf lost at the fair.
    There is the bird that the box lost at the fair.
    There is the bird that the wolf lost at the fair.
    He was very confused from the excitement.

20. That is the ball that the strawberry saw at the beach.
    That is the ball that the bird saw at the beach.
    That is the duck that the strawberry saw at the beach.
    That is the duck that the bird saw at the beach.
    He was too excited to even speak.

Appendix D

Filler Sentence Pairs

1. The boy swam with the man at the lake.
   The man found it very cold.
2. The man helped the nurse at school.
   The nurse was happy for the extra help.
3. The nurse waved at the doctor in the living room.
   The nurse thought the doctor was really friendly.
4. The doctor danced with the spider in the treehouse.
   The spider wanted to show off his fancy dance moves.
5. The spider smiled at the bee in the bathroom.
   The bee wanted them to become friends.
6. The bee cooked with the lady in the kitchen.
   The lady wanted to bake an apple pie.
7. The lady clapped with the king at the park.
   The lady was excited to go down the slide.
8. The king fought with the queen in the yard.
   The king did not want to share his toys.
9. The queen played with the clown at the farm.
   The queen was good at climbing trees.
10. The clown laughed with the boy in the store.
    The clown told a funny joke about a banana.
11. The police man read with the baby near the forest.
    The baby liked the story.
12. The farmer pushed the witch outside.
    The witch had hurt the farmer's feelings.
13. The witch colored with the baby behind the house.
    The baby thought it was a nice day out.
14. The farmer walked with the cowboy by the mountain.
    The farmer decided not to wear a warm jacket.
15. The cowboy skated with the dinosaur at the mall.
    The cowboy was clumsy and fell down.
16. The dinosaur jumped with the girl into the pond.
    The dinosaur wanted to cool off from the heat.
17. The girl talked to the bug at the river.
    The girl had a special secret to share.
18. The bug ran with the butterfly down the road.
    The bug wanted to get to the ice-cream truck first.
19. The butterfly whistled with the tiger at the beach.
    The tiger was afraid to go into the deep water.
20. The tiger sang with the police man at the fair.
    The police man wanted to win a prize.

Appendix E

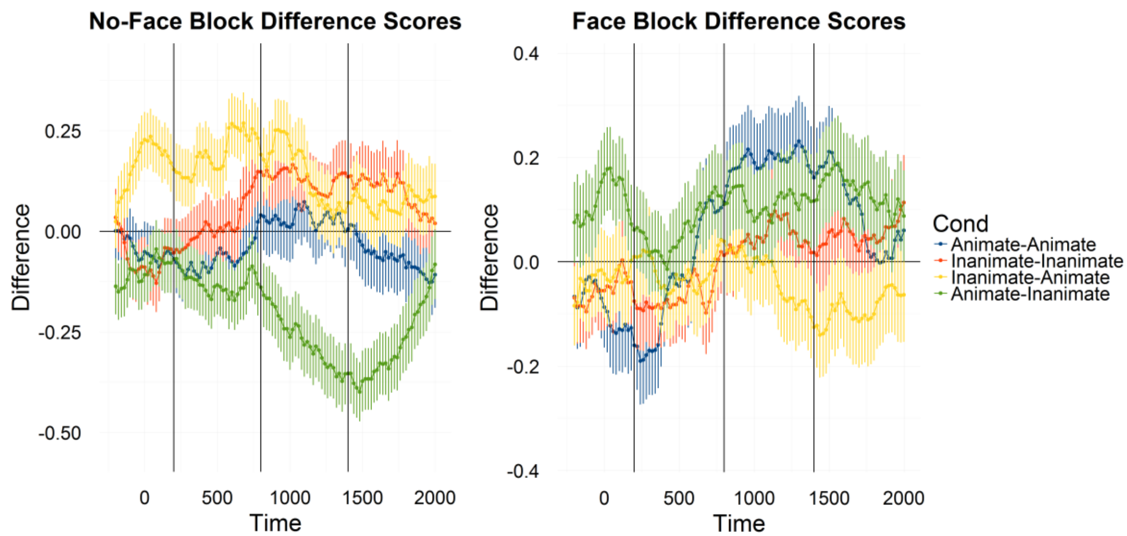

Figure E1. Child eye-tracking data by sentence type and image condition with error bars. Difference scores = looks to subject interest areas subtract looks to object interest areas. Vertical lines indicate 200, 800 and 1400 ms respectively.

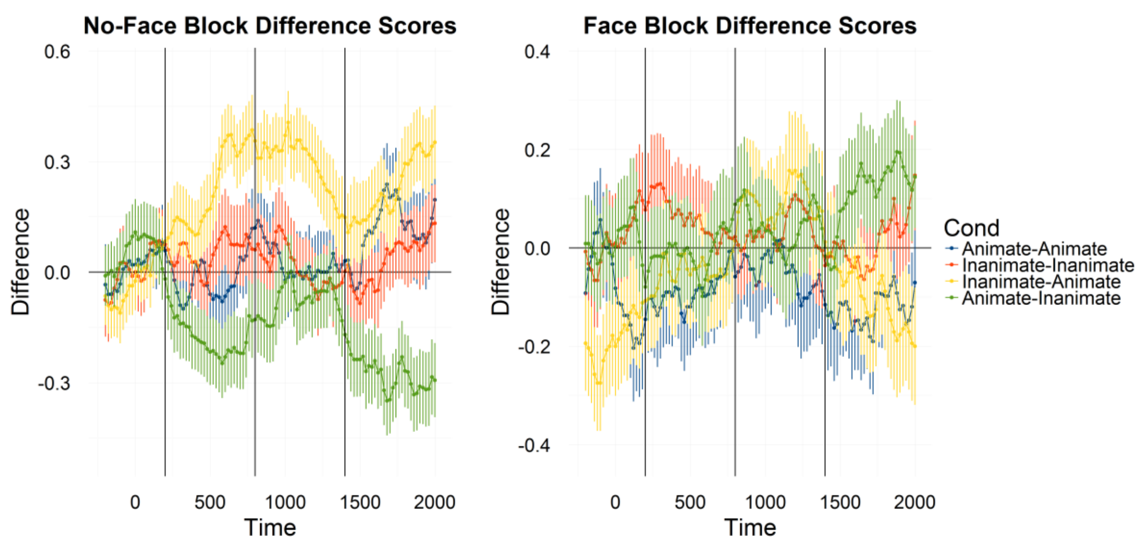

Figure E2. Adult eye-tracking data by sentence type and image condition with error bars. Difference scores = looks to subject interest areas subtract looks to object interest areas. Vertical lines indicate 200, 800 and 1400 ms respectively.