When Couches Have Eyes: The Effect of Visual Context on Children's Reference Processing

Rebecca Cooper-Cunningham [1], Monique Charest [1], Vincent Porretta [2] and Juhani Järvikivi [3]*

[1] Department of Communication Sciences and Disorders, University of Alberta, Edmonton, AB, Canada, [2] Department of Linguistics and Languages, McMaster University, Hamilton, ON, Canada, [3] Department of Linguistics, University of Alberta, Edmonton, AB, Canada

We examined the effects of semantic and visual cues to animacy on children's and adults' interpretation of ambiguous pronouns, using the visual world paradigm. Participants listened to sentences with object relative clauses that varied the animacy of potential referents, followed by test sentences beginning with a referentially ambiguous subject pronoun *he*. Participants viewed images of the referents, with semantically inanimate objects (e.g., a TV) shown with or without added facial features. Results from offline verbal report and online gaze data revealed consistent effects of both semantic animacy and visual context on pronoun resolution in both groups: There was a preference for semantically animate referents as antecedent, but this preference decreased or disappeared when semantically inanimate referents had facial features. The results indicate that the use of animacy as a linguistic cue is flexible and responsive to the visual context. They further suggest that like adults (Nieuwland and Van Berkum, 2006), 4-year-olds can already use fictional, here visual, context to adjust their online and offline language comprehension preferences.

Keywords: pronoun resolution, object relatives, animacy, visually situated language comprehension, child language processing, visual world eye tracking

INTRODUCTION

Animacy, or whether an entity is alive and volitional in the real world, is a semantic property with well-documented effects on language processing in both adults and children. The *semantic* animacy of objects, and their representation in various media, however, do not always align. It is commonplace in child-directed books and media for otherwise perfectly inanimate everyday objects, like the famous Brave Little Toaster (Disch, 1980; Kushner et al., 1987), Cars' Lightning McQueen (Anderson and Lasseter, 2006), or Beauty and the Beasts' Mrs. Teapot (Hahn et al., 1991), to have features such as eyes and mouths that signal animate agency, intentions, goals, and even personalities. Adult language comprehension quickly adapts to story contexts with inanimate objects acting as sentient agents (Nieuwland and Van Berkum, 2006), but the extent to which this is true for young children is not known. Of particular interest in the current study is whether and how visual cues to animacy (hereafter *visual* animacy) interact with semantic animacy to affect children's and adults' comprehension of reference. More specifically, we examine how visual and semantic animacy interact to influence processing of ambiguous pronouns.

In general, language users tend to take animate referents as representing syntactic subjects and semantic agents, especially if they occur in typical subject positions (see also Lowder and Gordon, 2015, for forces of nature as agents). Sentence subjects tend to be animate more often than inanimate (Clark and Begun, 1971). Thus, animate entities make better grammatical subjects, "doers" or "feelers," than inanimate ones. Within linguistic theory, thematic role relations rely heavily on the concept of animacy: only animate entities can be true agents and experiencers (Jackendoff, 1978), and prototypical agents are higher on the animacy hierarchy than prototypical patients, with properties corresponding to sentience, volition, and intentionality (Dowty, 1991; Kako, 2006). More generally, human referents are taken to be more prominent - and accessible - than animal referents, which, in turn, are taken as more prominent than inanimate referents (e.g., Comrie, 1989). In line with this expectation, corpus studies have shown that animate entities, in addition to being more frequently mentioned than inanimate entities (e.g., Givón, 1983), are much more likely to occur as sentence subjects and inanimate entities as objects (Clark and Begun, 1971; Dahl and Fraurud, 1996).

The match between such animacy-based expectations and the sentence structure (grammatical or thematic role assignment) affects sentence processing ease. ERP studies clearly demonstrate listeners' sensitivity when animacy expectations related to subjecthood or agency are violated (Weckerly and Kutas, 1999; Szewczyk and Schriefers, 2011; Nieuwland et al., 2013). Kuperberg et al. (2003), for example, showed that sentences with animacy violations (e.g., *For breakfast the eggs would only eat toast and jam*) elicited a strong P600 effect as compared to baseline sentences (e.g., *For breakfast the boys would only eat toast and jam*), whereas sentences with pragmatic violations (e.g., *For breakfast the boys would only bury toast and jam*) elicited only a weak (non-significant) N400 effect (see Kuperberg et al., 2007, for further evidence). More relevant to the present study, object relatives like *There is the snake/carrot that the bunny found* tend to show a processing cost compared to subject relatives (Gordon et al., 2001, 2002; Traxler et al., 2002, 2005). However, object relatives become easier to process when the head noun is inanimate than when it is animate (e.g., Ford, 1983; King and Just, 1991; Trueswell et al., 1994; Mak et al., 2002; Traxler et al., 2002; Clifton et al., 2003; Lowder and Gordon, 2014), with older adults showing this advantage more strongly than younger adults (DeDe, 2015).

Children are sensitive to animacy during language processing from an early age. Although English-speaking children rely strongly on word order to guide their interpretation of who-did-what-to-whom in standard Subject-Verb-Object sentences, animacy influences interpretations when word order is less available as a cue – such as for young children or non-canonical word orders (e.g., Bates et al., 1984; Thal and Flores, 2001). Animacy relationships also influence children's processing of relative clauses. Early studies suggested that processing of object relative clauses was of particular difficulty for children in comparison to subject relative clauses (Corrêa, 1995; Kidd and Bavin, 2002; Arnon, 2005). Indeed, in an act out study, accurate identification of the agent in centrally embedding object relatives like *The sheep that the pig pushed eats some grass*, was as

low as 55% for 5-year-old children, in contrast to identification rates of 83% for centrally embedded subject relatives, *The horse that jumped over the fence knocks down the chicken* (Corrêa, 1995). Differences in reliability of interpretation between subject and object relative clauses has been observed with sentences containing two animate referents (Dick et al., 2004) and more unusual sentences containing two inanimate referents (e.g., *The watch that had hugged the truck behind the kite was bright; The box that the kite had splashed behind the shoe was dry*; Montgomery et al., 2016). Comprehension differences between subject and object relatives disappear, however, when the animacy relations of the referents match with expectations for thematic and grammatical role assignment. For example, Kidd et al. (2007) examined children's abilities to imitate presentational-style right-branching relative clauses, which are the most common type of relative clause produced by children (Diessel and Tomasello, 2005). 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, e.g., *Here is the lady that helped the girl at school today*. Thus, effects of animacy on processing are wide ranging and, on the whole, consistent – both adults and children expect animate but not inanimate beings to do, think, feel, experience, and, importantly, serve as grammatical subjects and semantic agents.

However, as we obviously can and do imagine possible worlds where inanimate objects exhibit all these qualities, then what does it mean, from a language comprehension perspective, to be animate in a visually- or discourse-situated context? Nieuwland and Van Berkum (2006) demonstrated that linguistic representations of animacy can be highly flexible and rapidly responsive to the discourse context. They had adult participants listen to five-sentence stories with semantically anomalous characters (Experiment 1). In these stories, normally inanimate characters such as peanuts and yachts were treated as animate, in violation of normal 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 (Kutas and Federmeier, 2011, for a review), decreased (and was completely absent by the third mention), showing adaptation to the appearance of an inanimate agent. In Experiment 2, they created more elaborate stories, for example, a story about a peanut performing a song and a dance 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 that fit the discourse context than a semantically appropriate descriptor that would be in accordance with everyday experience. Thus, the context provided by the preceding story overruled default effects of animacy on processing. When debriefing with participants, Nieuwland and Van Berkum (2006) found that while listening to the test stories, the participants visualized the

inanimate nouns (e.g., the peanut) in anthropomorphic ways, such as cartoon-like, with faces and limbs.

The findings from Nieuwland and Van Berkum (2006) are consistent with the body of research showing that listeners take immediate advantage of the linguistic and non-linguistic context to shape their interpretations during comprehension (e.g., Trueswell et al., 1993; Spivey-Knowlton and Sedivy, 1995; Ferreira et al., 2013; Coco and Keller, 2015). It has been well-established that adult listeners can make immediate use of the visual context to constrain syntactic interpretation and resolve temporary syntactic (Tanenhaus et al., 1995; Spivey et al., 2002; Chambers et al., 2004; Ferreira et al., 2013), and referential (e.g., Altmann and Kamide, 1999; Sedivy et al., 1999; Knoeferle and Crocker, 2006) ambiguities in sentences. This literature indicates that the visual referential context and visual affordances incrementally inform sentence parsing strategies and referential choices (Chambers et al., 2004; Knoeferle and Crocker, 2007; Knoeferle et al., 2007). The results of Nieuwland and Van Berkum further show that discourse context can rapidly shape comprehenders' discourse representations and override normal effects of world knowledge on language processing (e.g., Hagoort et al., 2004; Hagoort and Van Berkum, 2007). The participants' reported visualizing strategies, moreover, suggest that visual input during comprehension could have effects on processing that may be similar to the kinds of effects observed for linguistic story context.

The present study asks whether visual context in the form of animate facial features influences processing of inanimate nouns in sentential contexts that have animate and inanimate nouns behaving in animate ways, and whether visual context of this kind might be sufficient to eliminate differences in referential processing expectations between animate noun and inanimate noun referents. This possibility is particularly germane for children's language processing. Children are consistently exposed to media portraying inanimate characters. It may seem self-evident to assume that children adapt to such storylines easily. However, when one considers the research surrounding children's more general use of visual context to guide linguistic interpretations, it is not clear how children's language processing adapts to these manipulations of visual features. Despite the evidence that adults use visual context to constrain language processing, children's ability to use visual information for the same purpose is less certain. Studies on the processing of temporarily ambiguous sentences suggests that 5-year-olds might not be sensitive to visual cues to the same extent as adults (Trueswell et al., 1999; Hurewitz et al., 2000; Snedeker and Trueswell, 2004; Weighall, 2008; Kidd et al., 2011). For example, when manipulating objects in the visual environment according to instructions like *Feel the frog with the feather*, where *with the feather* can either be analyzed as an instrument or as a modifier of the preceding noun phrase, *the frog*, children, unlike adults (e.g., Tanenhaus et al., 1995; cf. Ferreira et al., 2013), do not seem to benefit from the visual environment to constrain their syntactic parsing choices. Instead, they are influenced by sentence-internal, linguistic and distributional factors, such as whether the verb in question, (*e.g., feel vs. choose vs. tickle*), is followed with equal frequency (*feel*) or greater frequency by a

modifier (*choose*) or instrument (*tickle*) in everyday language use (Snedeker and Trueswell, 2004; Kidd et al., 2011).

Other studies, however, have shown that children are sensitive to visual contextual information when it serves referential processing rather than parsing (Weighall and Altmann, 2011; Zhang and Knoeferle, 2012; Huang and Snedeker, 2013; Hughes and Allen, 2013; Van Rij et al., 2016). Huang and Snedeker (2013), for example, found that 5-year-olds took into account visual scene information when interpreting the scalar adjectives "big" and "tall." They had 5-year-olds and adults listen to instructions such as *Point to the big* (vs. *small*) *coin* while they tracked participants' gaze to displays with one or two coins. In this task, both children and adults used the visual scene and looked to the correct item more quickly when the scene contained two possible referents compared to just one – even though children were delayed compared to adults. In relation to pronoun resolution, Van Rij et al. (2016) showed that young children's comprehension of object pronouns was dependent on whether the visual context showed a self- or other-directed action. In the same vein, Järvikivi and Pyykkönen-Klauck (2019) showed that 4-year-olds were more likely to choose as the referent of an ambiguous subject pronoun the character that was co-present in the visual context at the time of the pronoun than the one that left the screen before the pronoun onset, whereas co-presence did not change adult participants' pronoun resolution preferences. These studies suggest that young children in particular may be influenced by visual contextual information when it comes to referential processing, and especially when assigning reference to ambiguous pronouns.

As research demonstrates that animacy expectations in particular are contextually flexible and rapidly affected by discourse manipulation, it is of interest to ask whether visual contextual cues to animacy have a similar impact. This is especially relevant with respect to children's language processing: while their ability to recruit grammatical and semantic information for real-time language comprehension is in development, their daily experience is filled with story books and cartoons where inanimate objects play the role of sentient agents. Yet, whether children's use of contextual information in animacy processing is as flexible as adults' has not been previously examined. We simply assume that there is no conflict and that for both children and adults, visual cues to animacy would allow semantically inanimate referents to be processed more similarly to semantically animate ones when these behave in ways that contradict our perceived daily experience and the lexical meaning of the referents, thus aiding comprehension in accordance with the linguistic pragmatic context. Particularly relevant to children's language processing, inspecting the impact of visual vs. linguistic cues to animacy may show abilities to use visual context to guide processing that have been less evident with other aspects of language.

PRESENT STUDY

In the present study we used the visual world eye tracking paradigm and referent selection to examine how children's and adults' offline and online reference resolution preferences are affected by the animacy of the potential referent nouns, and

perhaps more interesting, by a visual context that suggests that an inanimate noun can behave in animate ways. We manipulated the visual animacy context by adding facial features (eyes and mouths) to cartoon images of inanimate objects. The neural basis of face recognition develops within the first 6 months of life, within which time faces also develop into a separate class of perceptual objects (Nelson, 2001). Prior research has established that the addition of eyes to inanimate objects serves as a visual cue to animacy even in infancy, affecting infants' categorization decisions (Welder and Graham, 2006; Anderson et al., 2018). Eye gaze signals speaker's attention to the listener (Langton et al., 2000), interlocutors attend closely to each other's faces (Argyle and Cook, 1976), and the effects of eye gaze are reflexive and present even when eye gaze is manipulated to not be an informative cue (e.g., Friesen and Kingstone, 1998). Eye-gaze has also been shown to affect pronoun resolution preferences when manipulated during (Nappa and Arnold, 2014) and preceding the pronoun (Hawthorne et al., 2016). As previous research suggests that linguistic story context can override normal assumptions about animacy and agentivity, we asked whether the same is the case with visual cues, and whether pronoun resolution in children and adults is affected by these cues to the same degree. A further question of interest was whether these effects appear immediately when visual-contextual information becomes available, or whether processing adapts to visual context more gradually.

Children and adults listened to sentence pairs that consisted of an object relative clause with a semantically animate or inanimate object noun plus a semantically animate or inanimate subject noun, followed by a sentence that began with the ambiguous pronoun *he* (e.g., *There is the snake/couch that the bunny/TV phoned. He was excited to go to the movie that night*). While listening, the participants viewed scenes depicting these characters, with semantically inanimate objects (couch, TV) depicted as either having eyes and mouths or not. We tracked their eye gaze to the pictures following the onset of the ambiguous pronoun *he*. We expected that animate nouns would be selected as the preferred referent for the pronoun *he* over inanimate nouns. However, we also expected this preference to be mitigated by the presence of facial features on images of the inanimate nouns.

Pronoun resolution provides an ideal context to examine questions about semantic and visual animacy for several reasons. English singular pronouns code explicitly for animacy, as in most cases, *he* and *she* will only refer to animate nouns. Pronouns refer to entities active in the listener's discourse representation and available as antecedents (Gernsbacher and Hargreaves, 1988; Gundel et al., 1993; Foraker and McElree, 2007). Their interpretation is inherently context-bound, making use of syntactic, semantic, and discourse information, as well as prior knowledge in building a coherent mental representation of the discourse event (e.g., Gernsbacher, 1989; Gordon et al., 1993; Almor, 1999; Pyykkönen and Järvikivi, 2010; Kehler and Rohde, 2013; Engelen et al., 2014; Van Dijk, 2014; Schumacher et al., 2017). Psycholinguistic studies have shown that pronoun processing preferences are facilitated by properties of the antecedents, in particular syntactic role and position, the subject or first-mentioned entity of the immediately preceding

clause/sentence often enjoying a more privileged status than other positions and roles (e.g., Gernsbacher et al., 1989; Gordon et al., 1993; Gundel et al., 1993; Arnold et al., 2000; Järvikivi et al., 2005; Kaiser and Trueswell, 2008). Pronouns are used to refer to human entities more frequently than non-human entities (Dahl and Fraurud, 1996). In line with this, animacy has been shown to modulate the selection of referring expressions when people continue sentences and stories: the likelihood of using a pronoun (vs. a full NP) is higher when continuations refer to animate entities than to inanimate entities (Fukumura and Van Gompel, 2011; Vogels et al., 2013, 2014). Moreover, recent processing studies using the visual world eye tracking paradigm have shown that children as young as 3–5 years of age already show many of these adult-like preferences, albeit often later in the time course (Song and Fisher, 2005, 2007; Arnold et al., 2007; Pyykkönen et al., 2010; Clackson et al., 2011; Järvikivi et al., 2014; Hartshorne et al., 2015), making pronoun resolution an appropriate tool to use. Bittner and Kuehnast (2012) have further shown that 5-year-old German and Bulgarian children resolve pronouns toward the subject more often when the subject is animate ("the monkey is hugging the dog") rather than inanimate ("the ball is touching the bear"). However, whether children of the same age benefit from the visual context during online pronoun resolution in general, and whether it incrementally modulates their use of animacy to determine subjecthood during processing, is not known.

Some research suggests that pronoun resolution effects may differ within and between sentences. Whether within and between sentence pronoun resolution is taken to rely on the same (Gordon and Hendrick, 1998) or different underlying mechanism (Miltsakaki, 2002), some research suggests that at least information structural effects, such as focusing, might affect anaphor resolution to a lesser extent within than between sentences (e.g., Colonna et al., 2012, 2015; Järvikivi et al., 2014). As animacy, the topic of the present study, has been shown to be a factor affecting referent prominence in 3-year old children's pronoun resolution (Pyykkönen et al., 2010), we chose to use intersentential rather than intrasentential materials. As our main question pertains to whether visual cues to animacy would affect referent prominence and thus pronoun resolution preferences, we wanted to give these cues a fair chance.

In the current study, the context sentences were presentational style right-branching object relative clauses. We chose presentational style relative clauses for the experimental stimuli due to the interpretability of the sentence frame for the age range of the children in this study (Kidd and Bavin, 2002; Diessel and Tomasello, 2005; Kidd et al., 2007). 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. Presentational-style right-branching relative clauses are the most common type of relative clause produced by children (Diessel and Tomasello, 2005).

Crucially, object relative clauses separate the grammatical subject and the first-mentioned noun. 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* is the first-mentioned noun. As discussed above, both first-mention and

subjecthood (or agenthood) have been shown to guide pronoun resolution preferences in adults and children (e.g., Frederiksen, 1981; Gernsbacher and Hargreaves, 1988; Crawley et al., 1990; Gordon et al., 1993; Carreiras et al., 1995; Järvikivi et al., 2005, 2014, 2017; Song and Fisher, 2005, 2007; Kaiser and Trueswell, 2008; Fukumura and Van Gompel, 2015; Hartshorne et al., 2015; Van Gompel and Järvikivi, 2016). By choosing a form that puts these two preferences into conflict, we increased the likelihood of uncovering effects of semantic and visual animacy, which may have otherwise been overshadowed by both subject preferences and first-mention preferences pointing toward the same noun.

METHODS

This study received research ethics approval from the University of Alberta Research Ethics Board, Project Name "Manipulating the Constraint of Animacy with Visual Context," No. Pro00049753.

Participants

Two groups of participants – adults and young children – participated in the experiment. Forty-three adult undergraduate university students were recruited through a participant database within the University of Alberta Linguistics Department and received course credit for participation. All adult participants completed the experiment at the Center for Comparative Psycholinguistics. Five of the adult participants were excluded from the analysis due to bilingual language background, and two were excluded due to technical difficulty with maintaining eye tracking. After exclusions, there were 36 participants in the adult group (18–52 years old, $M = 21.2$; 30 female). All included participants reported being monolingual English speakers with normal or corrected-to-normal vision, and normal hearing. Informed consent in writing was collected from adult participants and parents of child participants prior to participating in the experiment.

Forty-eight children aged 4;0–5;5 participated in the study. We recruited the children through community preschools and daycares, the Child and Adolescent Research Group database at the University of Alberta, and word of mouth. All included participants were monolingual English speakers, had normal corrected-to-normal vision, normal hearing, and typical language development according to parent report. Children reported as having a first language other than English, or who 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. Child participants received a t-shirt for participation.

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 et al., 2004) as an additional screen of language abilities. The Recalling Sentences subtest is standardized with a mean of 10 and a standard deviation of 3. No child scored more than 1 SD below the subtest mean (Range: 7–18, $M = 12.2$). One participant declined to participate in the sentence repetition task.

Given parent report of typical language development, this child's data were retained for analysis.

Participants were also excluded from the study if they did not correctly answer at least 14/20 comprehension questions from the filler items (described below), as we could not be certain that they had reliably attended to or understood the task. Four of the child participants were excluded for this reason. In addition, three 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 for 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
Linguistic Stimuli

The linguistic stimuli consisted of sentence pairs. The first sentence, the "context sentence" began with a demonstrative pronoun (*here, there, this, that*) and contained a presentational style right-branching object relative clause. The second sentence of each pair, the "test sentence" began with the ambiguous pronoun, *he* and made a general statement that was not specific to either of the referents. An example of a sentence pair is, *There is the snake that the bunny phoned at school. He was excited to go to the movie that night*. In the context sentence, *the bunny* is the subject of the relative clause and *the snake* is the object. The context sentence always ended with a prepositional phrase referring to a location in order to pull eye gaze away from the images of the subject and object before the pronoun in the test sentence was presented. Altogether 20 experimental and 20 filler pairs were created using 40 different transitive verbs (20 for experimental and 20 for filler items). The 20 verbs used in the experimental pairs were taken from Pyykkönen et al. (2010; see **Supplementary Material**). The subject (S) and object (O) noun of the relative clause were manipulated for semantic animacy (animate animal vs. inanimate object) in four conditions, resulting in four versions of the context sentence for each verb (*O Inanimate-S inanimate, O inanimate-S animate, O animate-S inanimate, and O animate-S animate*). **Table 1** presents one set of four context sentence versions plus the associated test sentence. The full list of experimental stimuli can be found in the **Supplementary Material**. All subject and object nouns had a minimum frequency of 50 occurrences per million words for the age range of 48–66 months in the ChildFreq database (Bååth, 2010).

The 20 filler items were similar to the experimental items, but did not include relative clauses or pronominal reference in the test sentence. The first sentence was a simple active declarative sentence ending with a mention of a location, similar to the experimental items. All subject and object characters in the filler sentences were animate, either human or animal. The second sentence directly referred to the subject or object of the previous sentence using a full noun phrase, for example, *The boy swam with the man at the lake. The man found it very cold*. The full list of filler pairs is available in the **Supplementary Material**.

The stimulus sentences were recorded by an adult female North American English speaker in a sound attenuated booth at a sampling frequency of 44.1 kHz. 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. A silent pause of 800 ms between sentences 1 and 2 was programmed into the experiment.

TABLE 1 | Example experimental materials.

	Head noun (object) animacy	Second noun (subject) animacy	Sentence
Context sentence	Inanimate	Inanimate	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	Inanimate	There is the *snake* that the <u>TV</u> phoned at school.
		Animate	There is the *snake* that the *bunny* phoned at school.
Test sentence			He was excited to go to the movie that night.

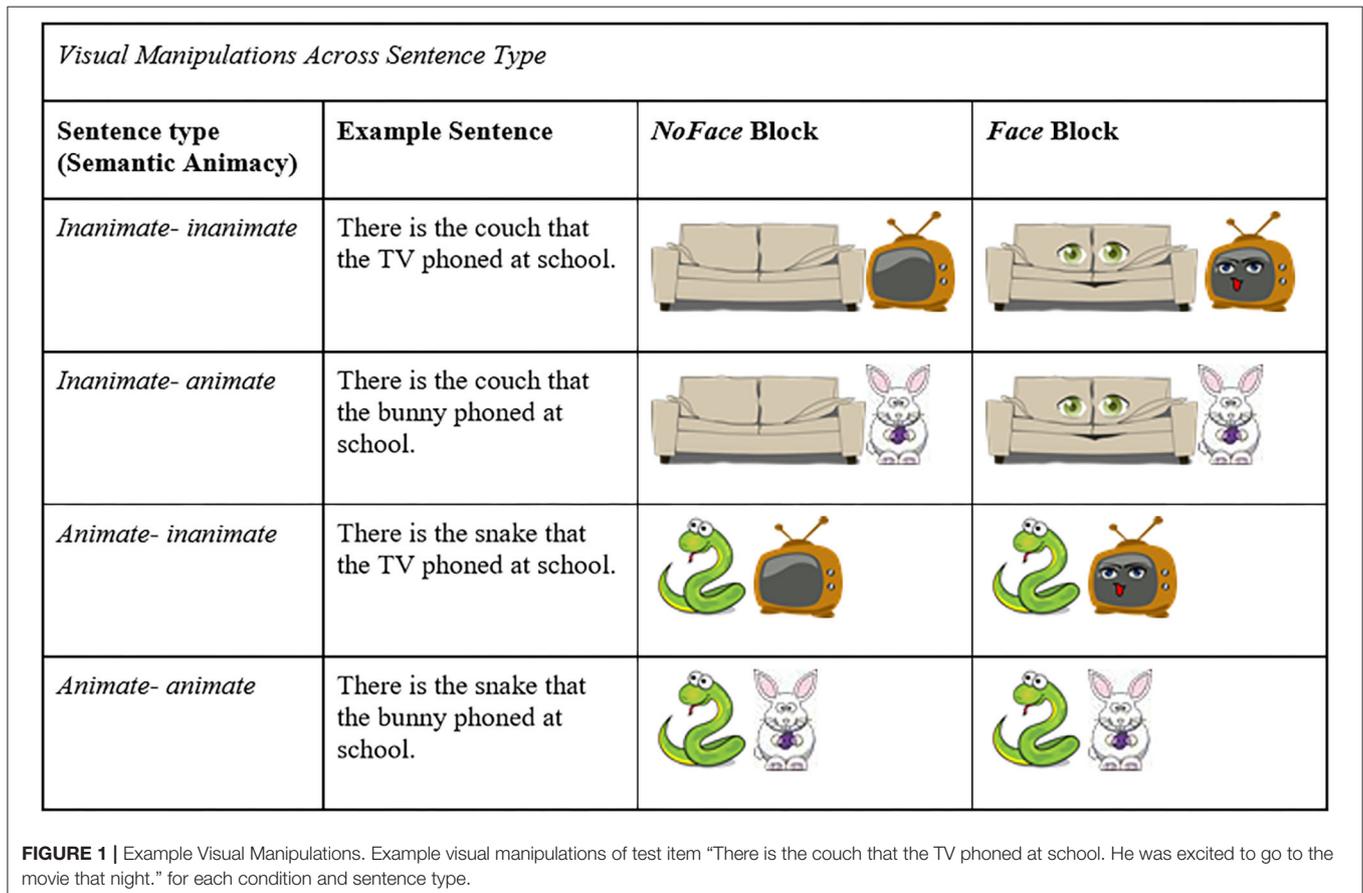*Inanimate referents underlined, animate referents in italics.*

Visual Stimuli

Each sentence pair was accompanied by stylized (cartoon) images of the subject and object characters (e.g., couch/snake and TV/bunny) and the location (e.g., school). We manipulated the visual animacy of the pictures of the semantically inanimate nouns (e.g., couch). This was done by taking the original picture of the inanimate object and adding eyes and a mouth. In one half of the trials, participants were shown the original images (visually inanimate: no face). In the other half of the trials they saw the manipulated image (visually animate: with a face). **Figure 1** provides an example of these manipulations. The full set of visual stimuli is available in the **Supplementary Material** (Experimental Images).

The position of the three images on the screen was counterbalanced between items. The images appeared in the top center, bottom left corner, and bottom right corner of the screen, equidistant from the center of the screen. Each image was placed inside an interest area 370 pixels wide by 330 pixels high. An example of the Visual World display where objects have been manipulated to be visually animate is available in **Supplementary Figure 1**.

Apparatus

All children were tested with an arm-mounted SR Research Eyelink 1000 Plus head-free eye-tracker, sampling at 500 Hz.

FIGURE 1 | Example Visual Manipulations. Example visual manipulations of test item "There is the couch that the TV phoned at school. He was excited to go to the movie that night." for each condition and sentence type.

Monitor size was 1,280 × 1,024 pixels. Participants' right eye was always tracked. The experiment was presented with SR Research Experiment Builder software (Version 1.10.1025). The linguistic stimuli were presented over table-top speakers. 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.

Seventeen of the adult participants were tested on the same eye-tracker as the child participants, and 19 of the adult participants were tested on a table-mounted SR Research Eyelink 1000 eye-tracker with a headrest, also sampling at 500 Hz. Screen size was 1,920 × 1,080 pixels.

Design and Procedure

Each participant completed two blocks of 20 trials, with 10 experimental and 10 filler trials per block. Visual animacy of the semantically inanimate nouns was blocked within the experiment and block order was counterbalanced between participants. That is, images with faces added on to the inanimate noun (*Face* Block) were presented in either Block 1 or Block 2. Images of the semantically animate nouns (e.g., bunny), did not vary between the two blocks. The order of the items was randomized within block, with the exception of the very first item of the experiment, which was always a filler item.

For both experimental and filler trials, stimulus items were assigned to Block 1 or 2 in a counterbalanced manner across participants. For the experimental items, each child was presented with one of the four possible versions created for each of the 20 verbs (i.e., *O Inanimate-S inanimate, O inanimate-S animate, O animate-S inanimate, and O animate-S animate*). For example, each participant was presented with only one of the four possible combinations of semantic animacy for the verb *phoned*, as presented in **Table 1**, with the visual features of the stimuli varied according to whether the trial appeared in the *Face* or *NoFace* Block. As there was only one sentence version created for each filler verb, the 20 filler items were the same for all participants.

Each of the four semantic animacy conditions was presented 5 times across the 20 experimental trials in the experiment for each participant. Within each block, 2 of the 4 semantic animacy conditions were presented twice, and 2 of the 4 semantic animacy conditions were presented three times for a total of 10 trials. The semantic animacy conditions that appeared twice in Block 1 appeared 3 times in Block 2, and vice-versa, to ensure 5 presentations of each semantic condition per participant. Six counterbalanced lists were created to account for all possible combinations of these distributions. In this way, we maintained equal counts for each semantic animacy condition across the full sample.

Participants were tested in quiet rooms/areas within their preschools and daycares, or at the Center for Comparative Psycholinguistics at the University of Alberta. Six adults and six children were assigned to each list, and block order was counterbalanced within those participants (i.e., three participants saw the visually inanimate block first, and three participants saw visually animate block first). The child participants were seated in front of a computer screen. The experimenter told the child

participants that they were going to hear some silly stories and see pictures on the screen. Each trial started with a cartoon star as a fixation point presented in the middle of the screen. When the child fixated the star, the experimenter started the trial. The visual display was presented for 1,000 ms, 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 comprehension question was, *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 in the exclusion criteria.

Responses were recorded in writing by the experimenter and coded with a key press as referring to the subject or object of the first sentence, or "other." Reasonable synonyms for the subject and object nouns were accepted as long as the referent was unambiguous, for example, "boot" for "shoe." Ambiguous responses (e.g., "the guy"), non-ambiguous responses that did not match either the subject or the object, and "I don't know" responses were coded as "Other." Between block 1 and 2, an optional movement break was offered to participants. After the experiment, the child participants completed the CELF-P2 Recalling Sentences subtest.

The procedure for adult participants was similar to the child participants, with some differences. 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 provided a verbal response.

RESULTS

We first report the data from the offline judgments of pronoun antecedent, followed by the online gaze data. For both data types, we report the effects of semantic and then visual animacy on pronoun resolution for the children first, followed by the adults.

Offline Response Data

Responses coded as "Other" (7.2% of child responses and 0.6% of adult responses) were removed prior to statistical analysis. The remaining responses were coded as referring to either the subject or the object of the first sentence. We analyzed the resulting binomial data (subject vs. object responses) with generalized linear mixed models using *lme4* in *R* (Bates et al., 2015), with the subject preference as the dependent variable. 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. We used backward elimination, starting with a model that included the main contrasts and interaction terms for sentence type and image condition, as well as the random intercepts for participants and items. The resulting model was tested against the model without the interaction term, and used function anova() to compare the models. We then tested for by-participant and by-item random slopes for sentence type and image condition.

To investigate a potential between-groups effect, we re-ran the offline response analysis on the combined dataset including a three-way interaction between sentence type, image condition, and group. Group was not statistically significant, neither in interaction $[X^2(7) = 4.4738, p > 0.5]$, nor as a main effect $[X^2(1) = 0.570, p > 0.1]$. Thus, to reduce model complexity and to investigate effects specific to each group, separate analyses were conducted.

Children

Figure 2A summarizes the distribution of children's verbal responses for each condition. Data are expressed as the raw difference in subject vs. object responses (raw number of object responses subtracted from subject responses). Visual inspection reveals an apparent trend for a subject advantage in the verbal responses of the children that appears to be stronger in the *NoFace* condition (dark gray bars). The exception to this trend is the *animate-inanimate* sentence type (e.g., *There is the snake that the TV phoned at school*). In this sentence type, there appears to be an object advantage when the subject (e.g., the TV) did not have a face on the image, and neither a subject nor an object advantage when the image on the subject contained facial features.

The verbal response data were analyzed with a binomial mixed-effects model including sentence type (four levels: *animate-animate, inanimate-inanimate, inanimate-animate,* and *animate-inanimate*) and image condition (two levels: *NoFace, Face*) as fixed predictors. Model comparison showed that adding the interaction between sentence type and image condition significantly improved model fit ($X^2 = 16.474, p < 0.001$). **Table 2** presents the summarized results. Positive estimates and *z*-values reflect a greater number of subject responses, while negative estimates and *z*-values reflect a greater number of object responses.
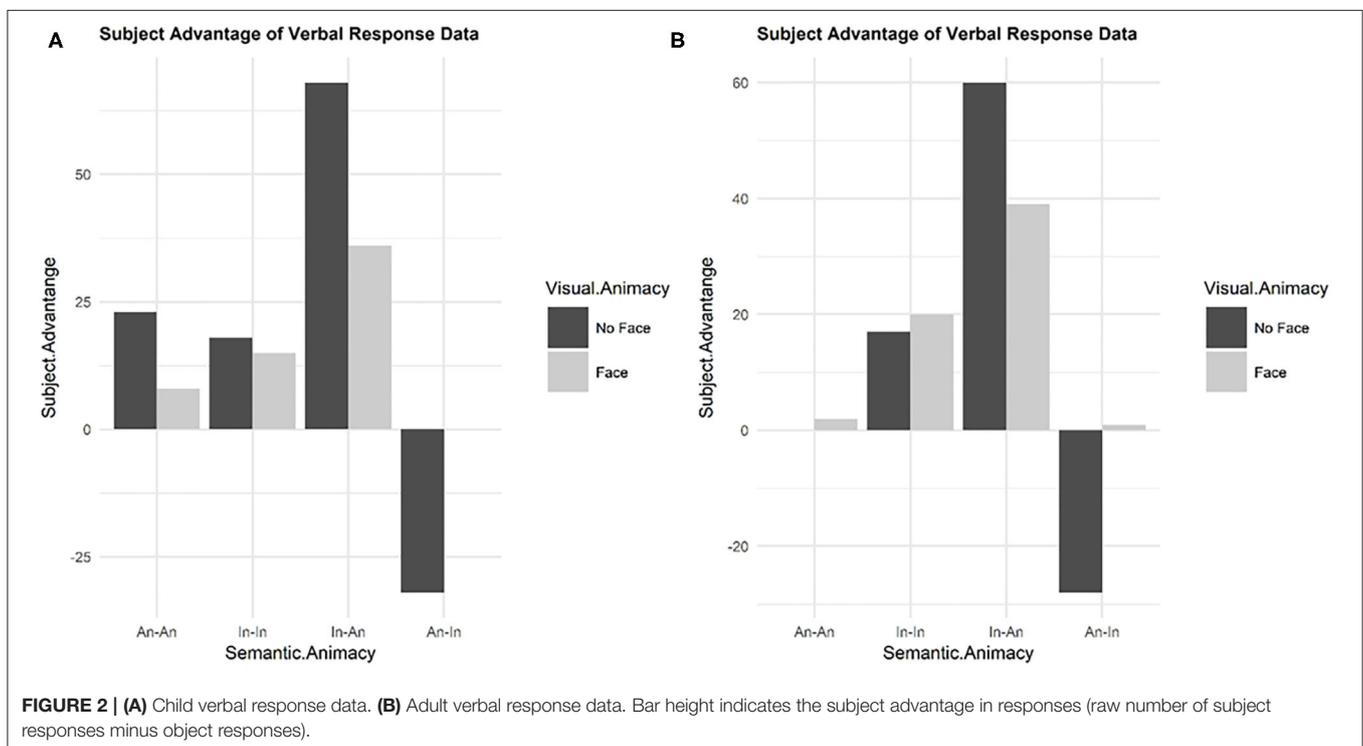
Semantic animacy

As the positive sign of the intercept shows, for the *animate-animate* sentences in the *NoFace* block, children

TABLE 2 | Child verbal response results: Mixed-effects modeling of verbal response to comprehension questions indicating subject noun vs. object noun interpretation of ambiguous pronoun.

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

The reference level was the animate-animate sentence type in the NoFace block.
*p < 0.05. **p < 0.01. ***p < 0.001.

FIGURE 2 | (A) Child verbal response data. **(B)** Adult verbal response data. Bar height indicates the subject advantage in responses (raw number of subject responses minus object responses).

showed a significant preference for the subject as antecedent in their verbal responses. Both the *inanimate-animate* and *animate-inanimate* sentence types were significantly different from the *animate-animate* intercept in the *NoFace* block. These differences are illustrated in **Figure 2A**. In comparison to the *animate-animate* sentences (e.g., *There is the snake that the bunny…*), the *inanimate-animate (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 *animate-inanimate* 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 suggests that the semantic animacy of the nouns affected the offline resolution of pronouns. Children were more likely to consider animate nouns as pronoun referents in their verbal responses.

Visual animacy

To investigate the interaction between sentence type and image condition, we ran pairwise comparisons between the *NoFace* and *Face* conditions for each sentence type, using package *emmeans* in *R* (Lenth, 2018). **Table 3** presents the results.

For both *inanimate-animate* (e.g., *…couch that the bunny…*) and *animate-inanimate* (e.g., *…snake that the TV…*) sentence types, adding faces significantly affected the offline responses to pronoun resolution. For *inanimate-animate* sentences, 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 *animate-inanimate* sentences, 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.

Adults

The verbal responses to the comprehension questions for adults were analyzed in the same manner as the child data. **Figure 2B** summarizes the distribution of adults' verbal responses for each condition. Visual inspection reveals similar trends in the adult data to that of the children. There remains a trend toward a subject advantage in the verbal responses. However, in contrast to

the children, there does not appear a subject or object advantage in the *animate-animate* sentence type. Similarly to the children, *animate-inanimate* sentence type suggests an object advantage when the subject (e.g., the TV) did not have a face on the image, and there was neither a subject or object advantage when the image on the subject contained facial features.

The verbal response data were again analyzed with a binomial mixed-model including sentence type and image condition as fixed predictors. Model comparison [function anova()] showed that adding the interaction term significantly improved model fit ($X^2 = 9.933$, $p = 0.019$). **Table 4** presents the summarized results.

Semantic animacy

In contrast to the children's verbal responses, the adult participants showed no significant preference for object or subject antecedents in the reference condition, *animate-animate NoFace*. Similar to the child verbal response data, in the *No Face* condition, *inanimate-animate* (e.g., "…*couch that the bunny…*") sentences showed significant differences to the intercept, illustrating that there is an effect of semantic animacy on pronoun resolution when measured offline through a verbal response (**Table 4**). There was an increase in participants indicating the animate subject (the bunny) as the antecedent in comparison to the *animate-animate* sentence type. Despite the visual trend in **Figure 2B** suggesting that participants indicated that the animate object (the snake) was the antecedent more often than in the *animate-animate* sentence type, this difference was not significant.

Visual animacy

The same pairwise comparisons of the verbal responses were completed for the adult group to explore the differences between image condition within the individual sentence types. **Table 5** presents the results.

Again, both *inanimate-animate* (e.g., "…*couch that the bunny…*") and *animate-inanimate* (e.g., "…*snake that the TV…*") sentence types showed significant differences in pronoun resolution with the addition of faces. For *inanimate-animate* sentences, the subject preference, and for *animate-inanimate*

TABLE 3 | Child verbal response results: Pairwise contrasts, comparing No Face and Face conditions for each sentence type.

Sentence Type	Estimate	SE	z-value	p-value
Animate - animate	0.378	0.320	1.180	0.278
Inanimate - inanimate	0.122	0.328	0.373	0.709
Inanimate - animate	1.266	0.427	2.963	0.003**
Animate - inanimate	−0.777	0.322	−2.412	0.016*

*p < 0.05. **p < 0.01.

TABLE 4 | Adult verbal response results: Mixed-effects modeling of verbal response to comprehension questions indicating subject noun vs. object noun interpretation of ambiguous pronoun.

Predictor	Estimate	SE	z-value	p-value
Intercept	−0.093	0.292	−0.319	0.750
Inanimate-inanimate	0.595	0.336	1.774	0.076
Inanimate-animate	1.965	0.387	5.082	3.74e−07***
Animate-inanimate	−0.637	0.338	−1.891	0.059
Face	0.159	0.325	0.489	0.625
Inanimate-inanimate: Face	−0.122	0.467	−0.261	0.7940
Inanimate-animate: Face	−1.009	0.514	−1.965	0.0495*
Animate-inanimate: Face	0.603	0.465	0.297	0.195

The reference level was the animate-animate sentence type in the NoFace block.
*p < 0.05. ***p < 0.001.

TABLE 5 | Adult verbal response results: Pairwise contrasts, comparing No Face, and Face conditions for each sentence type.

Sentence Type	Estimate	SE	z-value	p-value
Animate - animate	−0.159	0.325	−0.489	0.625
Inanimate - inanimate	−0.037	0.335	−0.111	0.912
Inanimate - animate	0.850	0.395	2.152	0.031*
Animate - inanimate	−0.762	0.333	−2.286	0.022*

*p < 0.05.

sentences the object preference, are decreased in strength significantly when faces were added, indicating that when the inanimate noun within a given sentence has animate visual characteristics, it is more likely to be considered as a pronoun referent.

Eye Tracking Data

The sample data were exported using SR Research Data Viewer (v 1.11.900), relative to the onset of the pronoun. Further processing of the data was done using the R package *VWPre* (Version 0.9.6; Porretta et al., 2016). The data were converted to proportion of samples falling within and outside each of the predefined interest areas in 20 ms windows (10 data points per 20 ms window). These proportions were then converted to empirical logits using 10 samples per bin and a constant of 0.5 (see Barr, 2008). Of interest here is the preference in looking behavior to the subject over the object of the relative clause. Therefore, the difference between looks to subject and looks to object was calculated, which we henceforth refer to as subject advantage. As it takes ~200 ms before eye gaze begins to reflect processing of an acoustic signal (Fischer, 1992; Matin et al., 1993), we used 200–2,000 ms post pronoun onset as the window for the subsequent analysis.

Model Fitting

Visual world eye tracking data are inherently a time series and the effect over time is typically non-linear. Of interest in visual world studies is how the time course is influenced by other variables (e.g., animacy of the objects on screen). We therefore used generalized additive mixed modeling (GAMM) (Hastie and Tibshirani, 1990; Wood, 2006; Baayen et al., 2017; Porretta et al., 2018). GAMM does not assume a linear relationship between continuous predictors (e.g., time) and the response variable (subject preference), and it allows us to model the time course of the effects without the need to aggregate over a series of shorter time windows. Additionally, GAMM also allows for the control of autocorrelation in the time series data (Baayen et al., 2018). Autocorrelation refers to the correlation between data points in a time series; a measurement at time point t is correlated to differing degrees with a measurement at time point t-i, depending on the lag. The presence of autocorrelation can lead to overconfidence of the model estimates. GAMM models were fitted in R using the package *mgcv* (Wood, 2016).

For both the child data and the adult data separately, a model was fitted for the response variable, subject advantage (see above). The model was fitted to the data using a backward
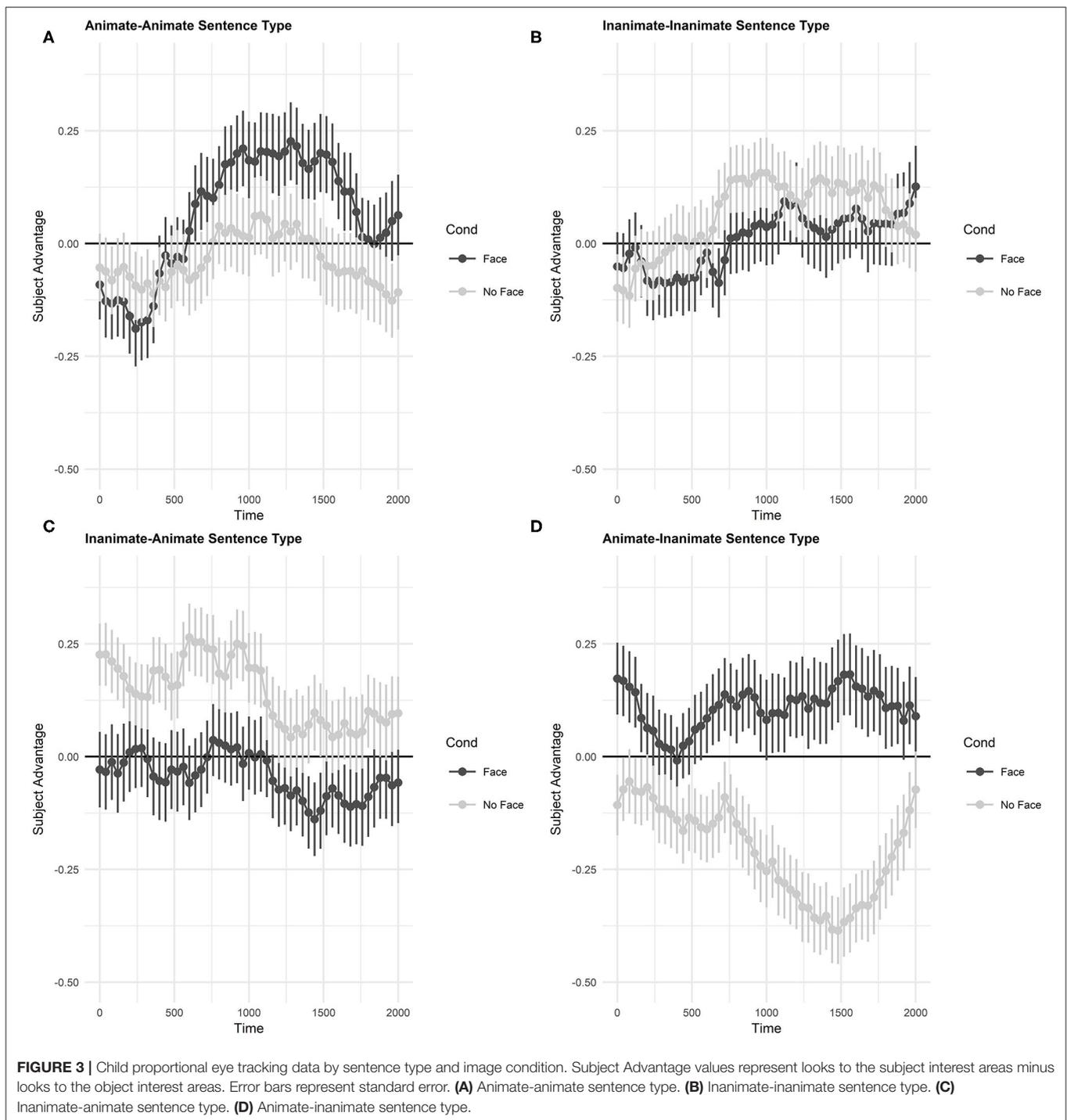
step-wise procedure (see Zuur et al., 2009) in order to evaluate the contribution of each predictor variable. First, we fitted a full model, that is, a model with all the predictors and interactions of interest, including random effects. Second, autocorrelation of the error was estimated from the model. The model was refitted with this parameter in order to adjust the confidence of the estimates. Third, we evaluated the contribution of the individual predictors in the model. For this, two criteria were used: the p-value of the term (indicating whether a given effect is not zero) and Maximum Likelihood (ML) score comparison between model variants (indicating whether the inclusion of the predictor improved the fit of the model). This process was done iteratively until the model contained only predictors that were statistically significant and contributed to the model fit.

For the child and adult models the following input variables were considered: animacy-by-face condition (the combination of animacy and face conditions of the visual scene), time, trial, subject, item, and event (the combination of subject and trial, indexing each unique time-series). The random structure consisted of random intercepts for events, factor smooths for time by subjects, and factor smooths for time by items. Factor smooths allow for the shape of the average time-course to vary by subject and item, as the time-course of the preference may vary for each. Random intercepts for event allow each unique time-course to have its own intercept in the model. For trial, a non-linear functional relation with the response variable was allowed for using a smooth function (Wood, 2006). For the interaction between time and animacy-by-face, a non-linear functional relation was allowed for using a smooth interaction (Wood, 2006). This interaction fits a curve for time for each level of animacy-by-face. In fitting the models (using the procedure mentioned above), only trial was removed for both the child and the adult data.

To investigate a potential between-groups effect, we ran the eye-tracking analysis on the combined dataset including an interaction between Manipulation and Group. Group was not statistically significant, neither in interaction [$X^2(7.0) = 4.243$, $p > 0.1$], nor as a main effect [$X^2(1.0) = 0.113$, $p > 0.5$]. Thus, to reduce model complexity and to investigate effects specific to each group, separate analyses were conducted. Below we present first the child data and model, followed by the adult data and model.

Children

Figure 3 displays the subject advantage scores for the eye tracking data, calculated by subtracting the proportion of looks to the object from the proportion of looks to the subject, and presents the gaze data from the *Face* and *NoFace* conditions, separately for each sentence type. The x-axis displays from the onset of the pronoun in the test sentence. Positive values indicate more looks to the subject image, while negative values indicate more looks to the object image. The dark gray lines in the figure depict the sentences presented in the *No Face* block and the light gray lines the sentences presented in the *Face* block. Thus, for semantically inanimate antecedent objects (e.g., sofa, phone) these lines show the difference in looks between when no eyes or mouth were added (dark gray lines) and with eyes and mouths added (light

FIGURE 3 | Child proportional eye tracking data by sentence type and image condition. Subject Advantage values represent looks to the subject interest areas minus looks to the object interest areas. Error bars represent standard error. **(A)** Animate-animate sentence type. **(B)** Inanimate-inanimate sentence type. **(C)** Inanimate-animate sentence type. **(D)** Animate-inanimate sentence type.

gray lines). For animate antecedents (snake, bunny), these lines depict the difference between whether they were presented in the *No Face* block (dark gray lines) or *Face* block (light gray lines). Note, in the case of the Animate_Animate sentences, with two semantically animate antecedents, the only difference is the block they were presented in.

Visual inspection of the data from the *NoFace* condition (dark gray lines) suggests that, in the absence of visual modification,

semantic animacy of the referents had a strong effect on pronoun resolution: There is a notable subject preference for *inanimate-animate* sentences which starts rising immediately after pronoun onset. In contrast, *animate-inanimate* sentences show a strong preference for object antecedents beginning around 750 ms. Visual inspection further suggests that there was no reliable subject or object preference for either the *animate-animate* or *inanimate-inanimate* sentence types, which suggests that there

was no systematic preference for either the subject or object noun in *animate-animate NoFace* sentences (dark gray line is not significantly different from the zero-line representing 50% or chance).

With the addition of facial features to the inanimate images (*Face* condition, light gray lines) the subject preference for *inanimate-animate* sentences, and the object preference for *animate-inanimate* sentences both decrease. That is, the subject advantage becomes closer to zero, and looking patterns appear more similar to the *animate-animate* and *inanimate-inanimate* sentences.

The results of the child model are reported in the **Supplementary Material**. **Supplementary Table 1** contains a summary of the model output. Part A of the table reports parametric coefficients related to adjustments to the intercept for each condition. Part B reports smooth terms related to time curves for each condition as well as the random effects.

Importantly, two of the curves for time were statistically different from zero, namely those for Animate_Animate_Face ($F = 3.9894$, $p = 0.003$) and Animate_Inanimate_No_Face ($F = 5.0748$, $p < 0.0001$). Additionally, as indicated by an EDF > 1 (3.072 and 4.474, respectively), these time curves were estimated to be non-linear. The *p*-value associated with each of these curves only indicates that the curve is not a zero line. In order to assess possible significant differences between the conditional curves predicted by the model (i.e., smooths for Time), visual inspection is needed, thus, it was necessary to calculate difference curves between the conditions. This was done using the R package *itsadug* (Van Rij et al., 2015). Below we report these separately for the effect of semantic animacy and visual animacy.

Semantic animacy

Figure 4 presents two difference curves related to the effect of semantic animacy. Each curve represents the difference between two model-predicted time curves and are shown with 99% confidence bands. The vertical lines indicate the statistically significant portion of the curve (i.e., the portion for which zero is not included in the confidence bands). The left panel displays the difference between Inanimate_Animate and Animate_Animate in the No Face condition. As can be seen, there were significantly more looks to the subject after the Inanimate_Animate than the Animate_Animate sentences during the whole analyzed time course until about 1,450 ms from the onset of the pronoun. The right panel displays the difference between Animate_Inanimate and Animate_Animate in the No Face condition. This curve shows that between 950 and 1,750 ms from the pronoun onset, there were significantly more looks to the object, or fewer looks to the subject, in the Animate_Inanimate condition as compared to the Animate_Animate one.

Visual animacy

Figure 5 presents two difference curves related to the effect of visual animacy. Each curve is presented with 99% confidence bands and vertical lines indicating statistical significance. The left panel displays the difference between Face and No Face in the Inanimate_Animate condition. As can be seen, adding visual features of animacy to pictures accompanying the Inanimate_Animate sentences resulted in significantly fewer subject looks starting at ∼500 ms after pronoun onset and lasting until about 1,400 ms after. The right panel displays the difference between Face and No Face in the Animate_Inanimate condition. This curve shows that after about 400 ms from the pronoun onset
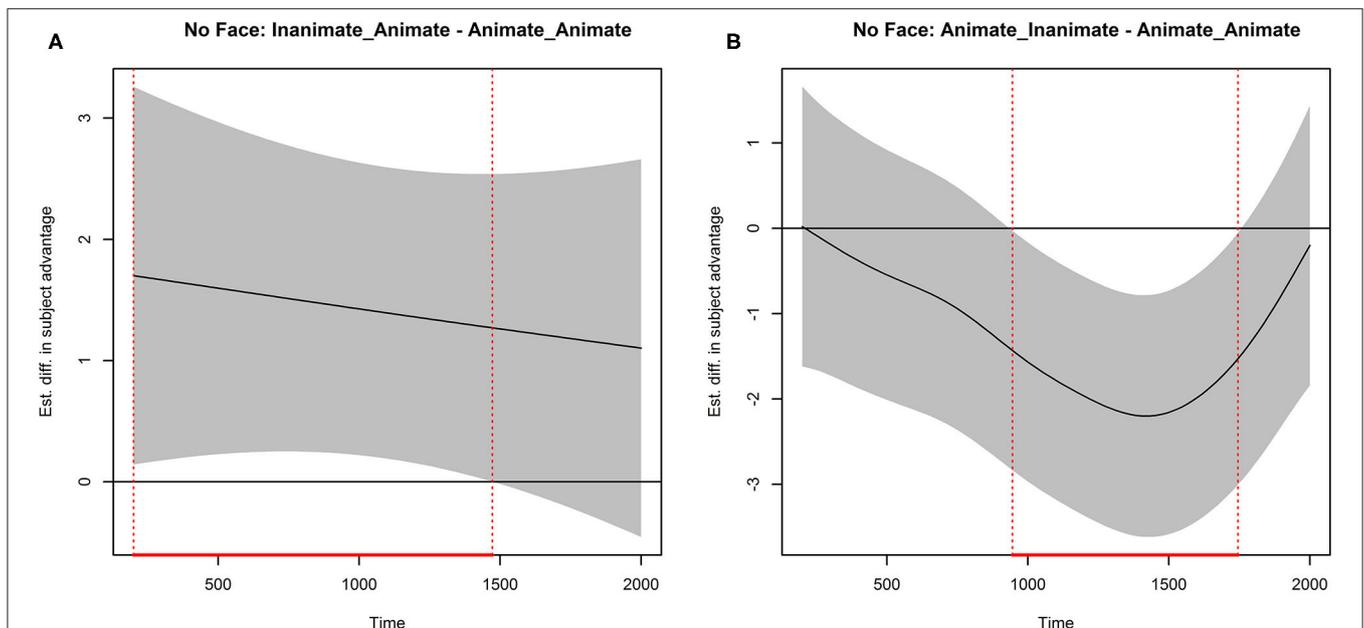
FIGURE 4 | Child data sentence type difference curves calculated from model-predicted curves with 99% confidence bands. Vertical lines indicate time points for which zero is not included in the confidence bands. **(A)** Inanimate-animate minus animate-animate (NoFace conditions). **(B)** Animate-inanimate minus animate-animate (NoFace conditions).

FIGURE 5 | Child data image condition difference curves calculated from model-predicted curves with 99% confidence bands. Vertical lines indicate time points for which zero is not included in the confidence bands. **(A)** Inanimate-animate Face minus NoFace conditions. **(B)** Animate-inanimate Face minus NoFace conditions.
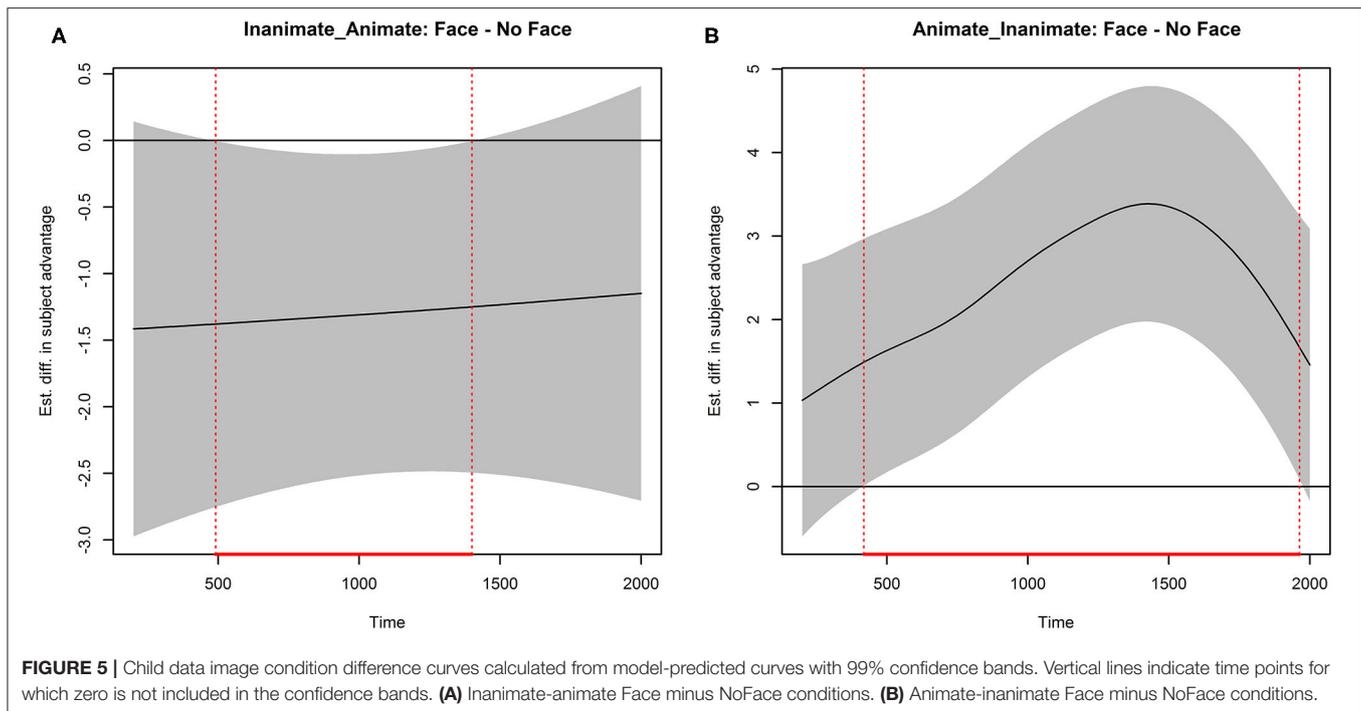
until at ~1,950 ms after, there were significantly more looks to the subject antecedent for Animate_Inanimate sentences with the added visual features of animacy than without.

Adults

Similar to the child data, **Figure 6** displays the subject advantage scores for the adult eye tracking data and presents the gaze data from the *Face* and *NoFace* conditions, separately for each sentence type. Positive values indicate more looks to the subject image, while negative values indicate more looks to the object image. Dark gray lines depict the sentences presented in the *No Face* block and light gray lines depict the sentences presented in the *Face* block.

Visual inspection of the data from the *NoFace* condition (dark gray lines) suggests that, similar to the child data, semantic animacy of the referents had a strong effect on pronoun resolution: The subject preference for *inanimate-animate* sentences starts rising about 250 ms after pronoun onset. As well, the object preference for *animate-inanimate* sentences begins around 500 ms. Also similar to the child data, visual inspection suggests no reliable subject or object preference for either the *animate-animate* or *inanimate-inanimate* sentence types, which suggests that there was no systematic preference for either the subject or object noun in *animate-animate NoFace* sentences (dark gray line is not significantly different from the zero line representing 50% or chance).
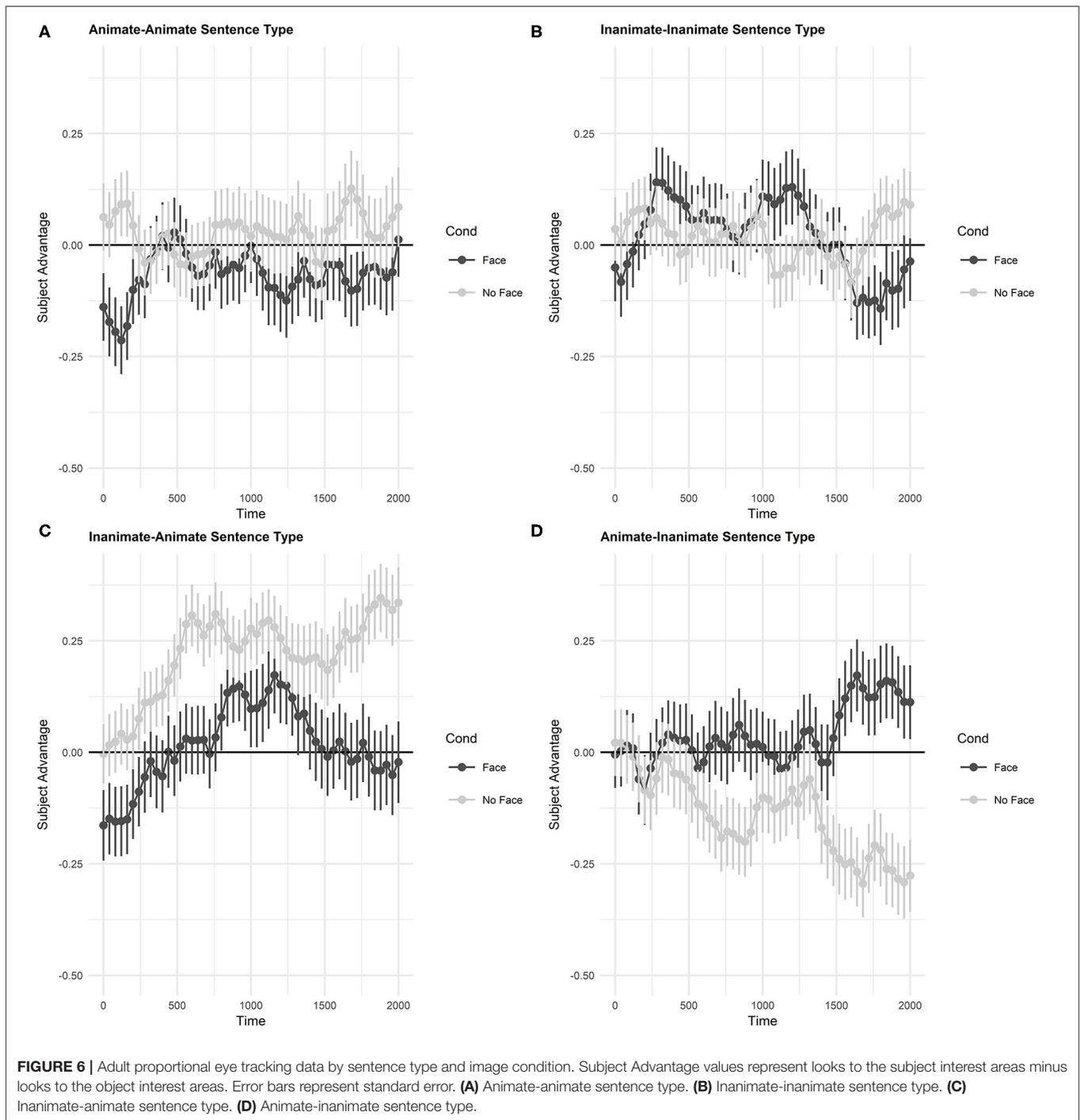
Adding facial features to the inanimate images (*Face* condition, light gray lines) had the same apparent effect in adult participants as in the child participants. The subject preference for *inanimate-animate* sentences, and the object preference for *animate-inanimate* sentences both decrease. The subject advantage becomes closer to zero, and looking patterns

appear more similar to the *animate-animate* and *inanimate-inanimate* sentences.

The full results of the adult model are reported in the **Supplementary Material**, **Supplementary Table 2** containing a summary of the model output (both parametric coefficients and smooth terms). Here, two of the curves for time were statistically different from zero, namely those for Animate_Inanimate_Face ($F = 4.3596$, $p = 0.037$) and Inanimate_Animate_No_Face ($F = 7.6981$, $p = 0.005$). Additionally, as indicated by EDF values of approximately one, these time curves were estimated to be linear. Again, the *p*-value associated with each of these curves only indicates that the curve is not a zero line. In order to assess possible significant differences between curves also accounting for differences in height (i.e., adjustments to the intercept), it was necessary to calculate difference curves between the conditions, as above. Below we report these separately for the effect of semantic animacy and visual animacy.

Semantic animacy

Figure 7 presents two difference curves related to the effect of semantic animacy. Each curve is presented with 99% confidence bands and vertical lines indicating statistical significance. The left panel displays the difference between Inanimate_Animate and Animate_Animate in the No Face condition. As with children, the adult data showed more looks to the subject for Inanimate_Animate than the Animate_Animate sentences. This effect was statistically significant starting at about 750 ms and continuing until 2,000 ms from the pronoun onset. The right panel displays the difference between Animate_Inanimate and Animate_Animate in the No Face condition. This curve shows significantly more object looks in Animate_Inanimate

FIGURE 6 | Adult proportional eye tracking data by sentence type and image condition. Subject Advantage values represent looks to the subject interest areas minus looks to the object interest areas. Error bars represent standard error. **(A)** Animate-animate sentence type. **(B)** Inanimate-inanimate sentence type. **(C)** Inanimate-animate sentence type. **(D)** Animate-inanimate sentence type.
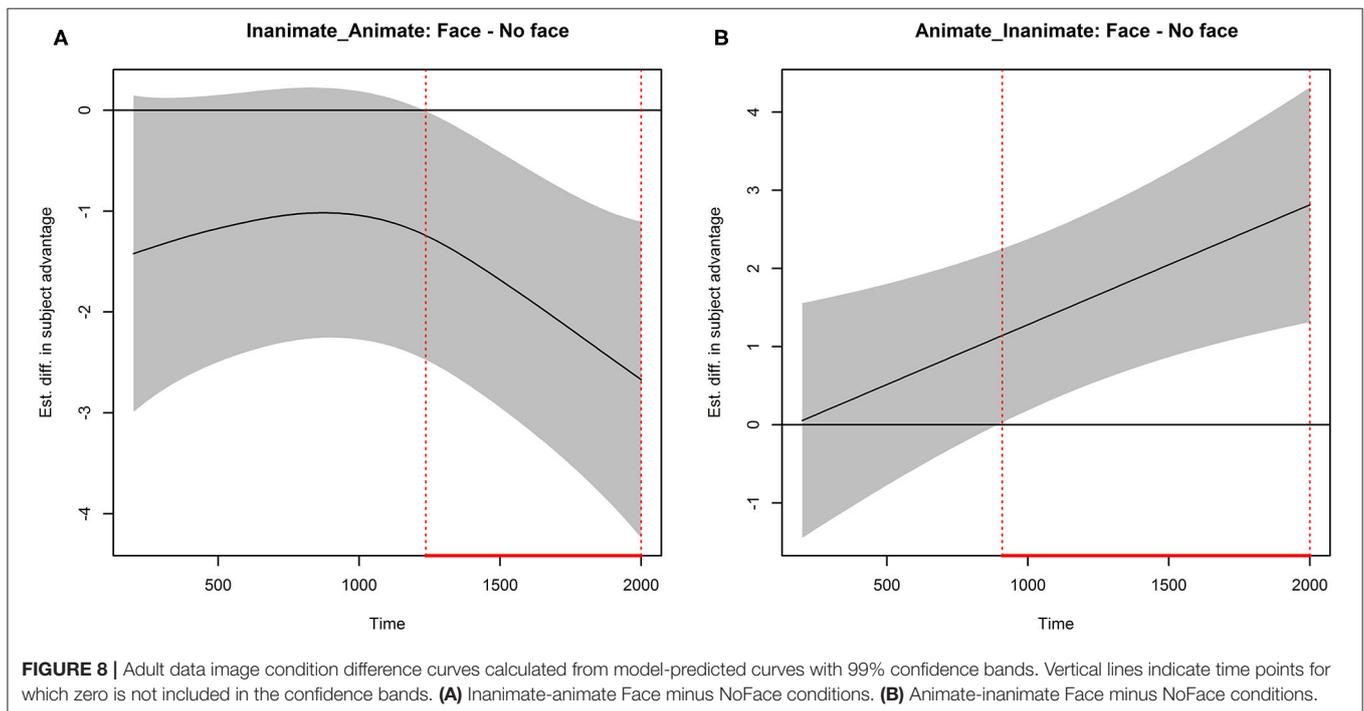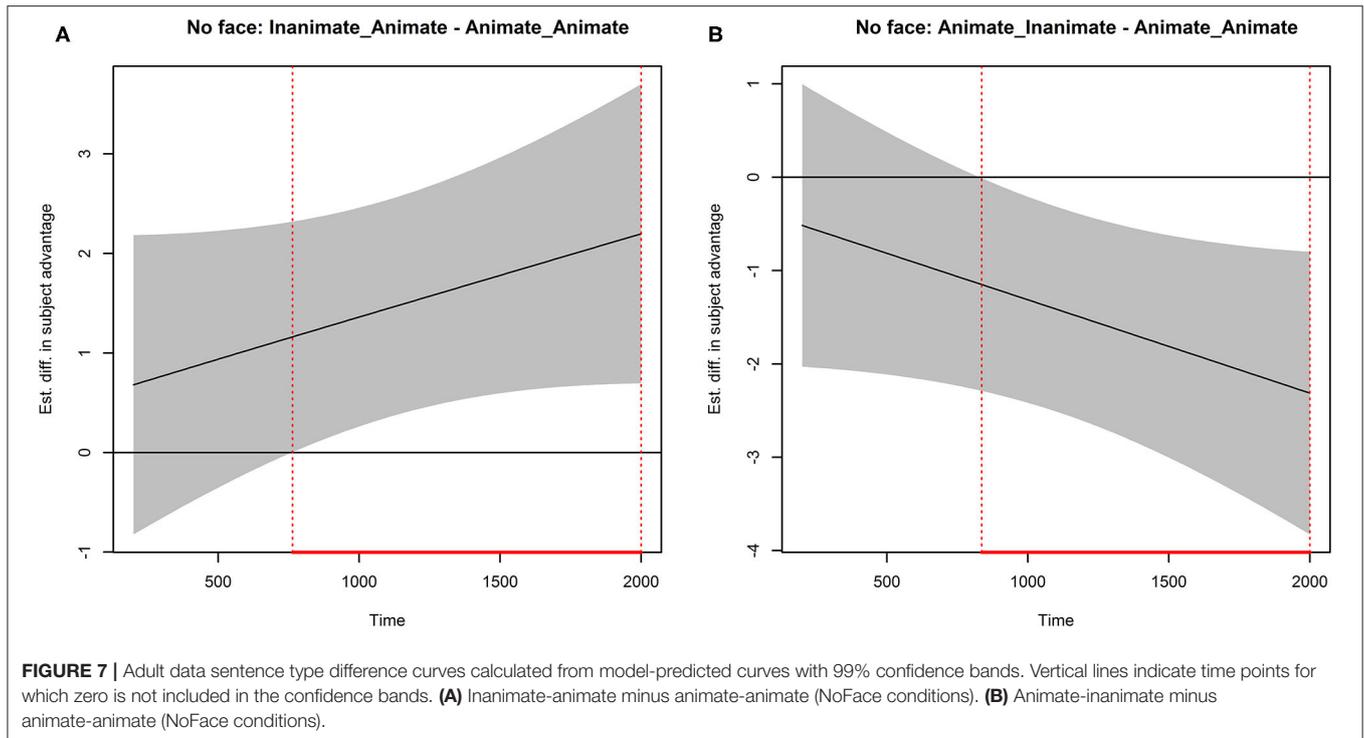
than Animate_Animate sentences. This effect was significant between 800 and 2,000 ms after pronoun onset.

Visual animacy

Figure 8 presents two difference curves related to the effect of visual animacy. Each curve is presented with 99% confidence bands and vertical lines indicating statistical significance. The left panel displays the difference between Face and No Face in the Inanimate_Animate condition. As the left panel indicates, adding

facial features to pictures accompanying the Inanimate_Animate sentences resulted in significantly increased looks to the object. This effect is visible from 1,200 ms after the pronoun onset and lasting until 2,000 ms after. The right panel displays the difference between Face and No Face in the Animate_Inanimate condition. This curve shows that after about 900 ms from the pronoun onset until 2,000 ms after, there were significantly more looks to the subject antecedent for Animate_Inanimate sentences with the added facial features than without.

FIGURE 7 | Adult data sentence type difference curves calculated from model-predicted curves with 99% confidence bands. Vertical lines indicate time points for which zero is not included in the confidence bands. **(A)** Inanimate-animate minus animate-animate (NoFace conditions). **(B)** Animate-inanimate minus animate-animate (NoFace conditions).

FIGURE 8 | Adult data image condition difference curves calculated from model-predicted curves with 99% confidence bands. Vertical lines indicate time points for which zero is not included in the confidence bands. **(A)** Inanimate-animate Face minus NoFace conditions. **(B)** Animate-inanimate Face minus NoFace conditions.

GENERAL DISCUSSION

The purpose of this study was to investigate how children would use semantic animacy and visual indicators of animacy during language comprehension, and to investigate whether similar effects are observed for children and adults. For both children and adults, we expected to see effects of both semantic and visual animacy on the proportion of looks to the subject noun vs. object noun during the presentation of the pronoun in the test sentence. This expectation was met. The results revealed that both semantic and visual animacy guided pronoun interpretation in offline, verbal response data and online, eye gaze data.

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., Trueswell et al., 1994; Clifton et al., 2003; Brandt et al., 2009; Lowder and Gordon, 2014). However, it has not been established how semantic animacy affects the online resolution of an animate, ambiguous pronoun, such as *he*, nor have the effects of using semantically anomalous nouns been established. Other semantic factors such as gender, verb transitivity, and agentivity have been shown to affect pronoun resolution in adults and children (Arnold et al., 2000, 2007; Rose, 2005; Pyykkönen et al., 2010; Schumacher et al., 2017). 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*. Moreover, like verb transitivity, animacy 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). 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. These two factors – providing distinguishing information and increasing saliency - suggest that animacy information should be used during pronoun resolution by both adults and children. In fact, this was found in offline and eye tracking results for both the children and adults.

The reference condition, *animate-animate* sentences, showed a significant subject preference in the child offline data, but not in the adult offline data. Even though in English, subjecthood and first-mention most often coincide, evidence from languages with less strict word order suggests that these two preferences are at least partly independent (e.g., Järvikivi et al., 2005). As the subject was the second, or most recent, mentioned antecedent in our stimuli, preference for the first-mentioned antecedent conflicted with preference for subject antecedent, predictably resulting in the pattern seen in the adult-data, when both the subject and object were animate. That children still showed a preference for the subject is in line with prior literature showing that the first-mention preference is less stable in children than in adults, at least until about 5 years of age, possibly longer and when it is observed, it is much weaker and appears later in time course than in adults (see Hartshorne et al., 2015, for an overview).

Inanimate-inanimate sentences never differed significantly from the animate-animate sentences, suggesting 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 toward the animate noun.

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 both the child and adult eye tracking data and offline verbal response data, inanimate nouns with faces were more likely to be considered as referents than the same nouns without, for both *inanimate-animate* and *animate-inanimate* sentence types.

This study differs from previous research on how visual context can affect language comprehension in children in two important ways. First, many previous studies investigated whether visual context affected children's syntactic parsing strategies (Trueswell et al., 1999; Hurewitz et al., 2000; Snedeker and Trueswell, 2004; Weighall and Altmann, 2011; Zhang and Knoeferle, 2012). In contrast, this study investigated whether visual context affected semantic interpretation of transitive sentences and of the subsequent anaphoric reference to the respective actors in the event. Other studies that focused on semantic interpretation 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 allowed us to observe the effects of various cues on language processing without requiring participants to formulate and revise hypotheses about sentence structure.

Semantic information is learned initially through connecting words to our real-world experiences that are accessed through visual and other sensory processes (Gleitman et al., 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 helps constrain 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 linguistic interpretation. 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 *animate-animate* and *inanimate-animate* sentences within the faces block.

In Nieuwland and Van Berkum (2006), the adaptation happened gradually. Our results suggest that when out-of-the-normal markers of animacy are presented visually, at least when these directly mark the actors in the story, participants adapt to it immediately. This is indicated by the absence of an effect of experimental trial or experimental block in the analyses.

However, even though block order did not significantly contribute to the model fit, if we inspect the effects of block order on children and adults, an interesting trend emerges. When participants received the Face block before the No Face block, their processing in the latter clearly reflects this fact. As **Supplementary Figure 1** shows, participants' subject preference

for *animate-animate* and *animate-inanimate* sentences in the No Face block mirrors that in the Face block when the latter precedes the former. However, this is not the case the other way around. This suggests that even when visually animate features on the inanimate objects go away in the second block, both children and adults continue to uphold the animacy assumptions as if they would be the same as in the first block. This suggests that the presence of visual cues in the first block changed how adults and children processed the sentence pairs even when these cues were no longer present. This raises the question of what would happen if linguistic story context were pitted against a visual context. Whether one or the other would take precedence, whether the effects would be the same for adults and children in that case, and whether the effects of adaptation would be the same for these two type of contexts, are interesting questions that deserve to be answered in future research (see Lee et al., 2017).

There are some limitations to this study that deserve a mention. First, both the child and adult participant groups were fairly homogenous in terms of socioeconomic status (SES), and further, most of the parents had a high educational background. 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). 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. Such findings would help distinguish if differences between adults and children are due to the relevance of the visual information to the individuals' lives, or if the differences are due to children's inability to inhibit the irrelevant visual information. Ours was also a fairly homogenous group in terms of age and language background. This leaves open when exactly children start showing adult-like sensitivity to animacy. Thus, further research might shed light at the age at which children start exhibiting adult-like sensitivity for animacy – semantic and visual – as a pronoun resolution cue; and, further, whether and the extent to which general language development (or experience) would predict their performance.

Second, coming back to the issue of inter- vs. intrasentential pronoun resolution, it might be that our use of between-sentence materials facilitated the effects of visual animacy, and a visual context might have less of an effect within sentence. However, these effects in both adults and children were relatively rapid, appearing well-within the time in which effects of subject-/agenthood have been shown in prior experiments with ambiguous pronouns in adults (e.g., Järvikivi et al., 2005; Clackson et al., 2011; Schumacher et al., 2017), and even earlier than what has been usually observed in children (Hartshorne et al., 2015, for an overview), suggesting that pronoun resolution processes were as immediate as could have been expected if the antecedent had been in the same sentence. This suggests that while the overall pattern of preferences could conceivably have been different had we opted for testing pronoun resolution within- rather than between sentences (but see Colonna et al., 2015), there is less reason to believe that the effects of visual animacy would have changed. However, as our aim was to see

if visual animacy would have an effect in the first place and when during processing, this is a question for future research to answer. Lastly, a secondary finding to this study, the early online first-mention preference and 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 preferences in both children and adults. Relative clauses offer a unique opportunity in English to separate the two constraints, first-mention and subjecthood, 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 how it develops in 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 found 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 timeframes in the eye-tracking data. This suggests that already at 4 years of age, children use the semantic and visual cues to animacy similarly to adults to guide their pronoun interpretation. All observed effects in this study suggest that animate nouns, whether semantically or visually so, are preferred over inanimate nouns during pronoun resolution. Thus, when faced with decisions about ambiguous pronouns, both adults and children still use semantic animacy to some degree separately from visual animacy.

The findings of this study suggest that, much like adults, already 4-year-old children are capable of using semantic and visual cues to animacy to aid subsequent pronoun resolution information during language comprehension, and that in both age groups, use of animacy as a linguistic cue is flexible and responsive to the visual context.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Alberta Research Ethics Office. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MC, RC-C, and JJ conceptualized and designed the study. RC-C collected the data. MC, RC-C, JJ, and VP analyzed, interpreted the data, wrote, and commented on the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2020.576236/full#supplementary-material

REFERENCES

Almor, A. (1999). Noun-phrase anaphors and focus: the informational load hypothesis. *Psychol. Rev.* 106, 748–765. doi: 10.1037/0033-295X.106.4.748

Altmann, G. T. M., and Kamide, Y. (1999). Incremental interpretation of verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1

Anderson, D. K., and Lasseter, J. (2006). *Cars. [Motion picture]*. Burbank, CA: Walt Disney Pictures & Pixar Animation Studios

Anderson, N., Meagher, K., Welder, A., and Graham, S. A. (2018). Animacy cues facilitate 10-month-olds' categorization of novel objects with similar insides. *PLoS ONE* 13:e0207800. doi: 10.1371/journal.pone.0207800

Argyle, M., and Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge, MA: Cambridge University Press

Arnold, J. E., Brown-Schmidt, S., and Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Lang. Cogn. Process* 22, 527–565. doi: 10.1080/01690960600845950

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., and Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition* 76, B13–B26. doi: 10.1016/S0010-0277(00)00073-1

Arnon, I. (2005). "Relative clause acquisition in Hebrew: Towards a processing-oriented account," In *Proceedings of the 29th Boston University Conference on Language Development*. (Somerville, MA: Cascadilla Press).

Bååth, R. (2010). ChildFreq: An Online Tool to Explore Word Frequencies in Child Language. LUCS Minor, 16, 1–6.

Baayen, H., Vasishth, S., Kliegl, R., and Bates, D. (2017). The cave of shadows: addressing the human factor with generalized additive mixed models. *J. Mem. Lang.* 94, 206–234. doi: 10.1016/j.jml.2016.11.006

Baayen, R. H., Rij, J., van Cat, C., and de Wood, S. N. (2018). "Autocorrelated errors in experimental data in the language sciences: some solutions offered by generalized additive mixed models," in *Mixed Effects Regression Models in Linguistics*. eds Speelman, D., Heylen, K., and Geeraerts, D. (Berlin: Springer), 49–69. doi: 10.1007/978-3-319-69830-4_4

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *J. Mem. Lang.* 59, 457–474. doi: 10.1016/j.jml.2007.09.002

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., and Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Dev.* 55, 341–354. doi: 10.2307/1129947

Bittner, D., and Kuehnast, M. (2012). Comprehension of intersentential pronouns in child German and child Bulgarian. *First Lang.* 32, 176–204. doi: 10.1177/0142723711403074

Brandt, S., Kidd, E., Lieven, E., and Tomasello, M. (2009). The discourse bases of relativization: an investigation of young German and English-speaking children's comprehension of relative clauses. *Cogn. Linguist.* 20, 539–570. doi: 10.1515/COGL.2009.024

Carreiras, M., Gernsbacher, M. A., and Villa, V. (1995). The advantage of first mention in Spanish. *Psychon. Bull. Rev.* 2, 124–129. doi: 10.3758/BF03214418

Chambers, C. G., Tanenhaus, M. K., and Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *J. Exp. Psychol.* 30:687. doi: 10.1037/0278-7393.30.3.687

Clackson, K., Felser, C., and Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: evidence from eye-movements during listening. *J. Mem. Lang.* 65, 128–144. doi: 10.1016/j.jml.2011.04.007

Clark, H. H., and Begun, J. S. (1971). Semantics of sentence subjects. *Lang. Speech* 14, 34–46. doi: 10.1177/002383097101400105

Clifton, C. Jr., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., and Rayner, K. (2003). The use of thematic role information in parsing: syntactic processing autonomy revisited. *J. Mem. Lang.* 49, 317–334. doi: 10.1016/S0749-596X(03)00070-6

Coco, M. I., and Keller, F. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *Quart. J. Exp. Psychol.* 68, 46–74. doi: 10.1080/17470218.2014.936475

Colonna, S., Schimke, S., and Hemforth, B. (2012). Information structure effects on anaphora resolution in German and French: a crosslinguistic study of pronoun resolution. *Linguistics* 50, 991–1013. doi: 10.1515/ling-2012-0031

Colonna, S., Schimke, S., and Hemforth, B. (2015). Different effects of focus in intra-and inter-sentential pronoun resolution in German. *Lang. Cogn. Neurosci.* 30, 1306–1325. doi: 10.1080/23273798.2015.1066510

Comrie, B. (1989). *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press

Cooper, R. (2016). *Cartoons and Comprehension: The Effect of Visual Context on Children's Sentence Processing. 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.

Corrêa, L. M. S. (1995). An alternative assessment of children's comprehension of relative clauses. *J. Psycholinguist. Res.* 24, 183–203. doi: 10.1007/BF02145355

Crawley, R. A., Stevenson, R. J., and Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *J. Psycholinguist. Res*, 19, 245–264. doi: 10.1007/BF01077259

Dahl, Ö., and Fraurud, K. (1996). "Animacy in grammar and discourse," in *Reference and Referent Accessibility*. eds T. Fretheim and J. K. Gundel (Amsterdam: John Benjamins), 47–64. doi: 10.1075/pbns.38.04dah

DeDe, G. (2015). Effects of animacy on processing relative clauses in older and younger adults. *Quart. J. Exp. Psychol.* 68, 487–498. doi: 10.1080/17470218.2014.956766

Dick, F., Wulfeck, B., Krupa-Kwiatkowski, M., and Bates, E. (2004). The development of complex sentence interpretation in typically developing children compared with children with specific language impairments or early unilateral focal lesions. *Dev. Sci.* 7, 360–377. doi: 10.1111/j.1467-7687.2004.00353.x

Diessel, H., and Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language* 81, 882–906. doi: 10.1353/lan.2005.0169

Disch, T. M. (1980). The brave little toaster. *The Magazine of Fantasy and Science Fiction* 59, 2. Cornwell, CT.

Dowty, D. R. (1991). Thematic proto-roles and argument selection. *Language* 67, 547–619. doi: 10.1353/lan.1991.0021

Engelen, J. A., Bouwmeester, S., de Bruin, A. B., and Zwaan, R. A. (2014). Eye movements reveal differences in children's referential processing

during narrative comprehension. *J. Exp. Child Psychol.* 118, 57–77. doi: 10.1016/j.jecp.2013.09.005

Ferreira, F., Foucart, A., and Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *J. Mem. Lang.* 69, 165–182. doi: 10.1016/j.jml.2013.06.001

Fischer, B. (1992). "accadic reaction time: implications for reading, dyslexia, and visual cognition," in *Eye movements and visual cognition, Springer Series in Neuropsychology.* eds K. Rayner (New York, NY: Springer), 31–45. doi: 10.1007/978-1-4612-2852-3_3

Foraker, S., and McElree, B. (2007). The role of prominence in pronoun resolution: active versus passive representations. *J. Mem. Lang.* 56, 357–383. doi: 10.1016/j.jml.2006.07.004

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *J. Verbal Learn. Verbal Behav.* 22, 203–218. doi: 10.1016/S0022-5371(83)90156-1

Frederiksen, J. R. (1981). Understanding anaphora: rules used by readers in assigning pronominal reference. *Discourse Process* 4, 323–347. doi: 10.1080/01638538109544525

Friesen, C. K., and Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon. Bull. Rev.* 5, 490–495. doi: 10.3758/BF03208827

Fukumura, K., and Van Gompel, R. P. (2011). The effect of animacy on the choice of referring expression. *Lang. Cogn. Process* 26, 1472–1504. doi: 10.1080/01690965.2010.506444

Fukumura, K., and Van Gompel, R. P. G. (2015). Effects of order of mention and grammatical role on anaphor resolution. *J. Exp. Psychol. Learn. Memory Cogn.* 41, 501–525. doi: 10.1037/xlm0000041

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition* 32, 99–156. doi: 10.1016/0010-0277(89)90001-2

Gernsbacher, M. A., and Hargreaves, D. J. (1988). Accessing sentence participants: the advantage of first mention. *J. Mem. Lang.* 27, 699–717. doi: 10.1016/0749-596X(88)90016-2

Gernsbacher, M. A., Hargreaves, D. J., and Beeman, M. (1989). Building and accessing clausal representations: the advantage of first mention versus the advantage of clause recency. *J. Mem. Lang.* 28, 735–755. doi: 10.1016/0749-596X(89)90006-5

Givón, T. (1983). "Topic continuity in discourse: An introduction," in *Topic Continuity in Discourse: A Quantitative Cross-Language Study.* ed T. Givón (Amsterdam: John Benjamins), 1–42. doi: 10.1075/tsl.3

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., and Trueswell, J. C. (2005). Hard words. *Lang. Learn. Dev.* 1, 23–64. doi: 10.1207/s15473341lld0101_4

Gordon, P. C., Grosz, B. J., and Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cogn. Sci.* 17, 311–347. doi: 10.1207/s15516709cog1703_1

Gordon, P. C., and Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cogn. Sci.* 22, 389–424. doi: 10.1207/s15516709cog2204_1

Gordon, P. C., Hendrick, R., and Johnson, M. (2001). Memory interference during language processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1411–1423. doi: 10.1037//0278-7393.27.6.1411

Gordon, P. C., Hendrick, R., and Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychol. Sci.* 13, 425–430. doi: 10.1111/1467-9280.00475

Gundel, J. K., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69, 274–307. doi: 10.2307/416535

Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441. doi: 10.1126/science.1095455

Hagoort, P., and Van Berkum, J. (2007). Beyond the sentence given. *Philos Trans R Soc B Biol Sci.* 362, 801–811. doi: 10.1098/rstb.2007.2089

Hahn, D., Trousdale, G., and Wise, K. (1991). *Beauty and the Beast [Motion picture].* United States of America: Walt Disney Pictures, Walt Disney Feature Animation, & Silver Screen Partners IV

Hartshorne, J. K., Nappa, R., and Snedeker, J. (2015). Development of the first-mention bias. *J. Child Lang.* 42, 423–446. doi: 10.1017/S030500091400075

Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized Additive Models. Monographs on Statistics and Applied Probability,* Vol. 43. London: Chapman & Hall

Hawthorne, K., Arnhold, A., Sullivan, E., and Järvikivi, J. (2016). "Social cues modulate cognitive status of discourse referents," in *Proceedings of the 38th Annual Conference of the Cognitive Science Society.* (Austin, TX: Cognitive Science Society), 562–567.

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Dev.* 74, 1368–1378. doi: 10.1111/1467-8624.00612

Huang, Y. T., and Snedeker, J. (2008). "Use of referential context in children's language processing," in *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (Austin, TX). 1380–1385.

Huang, Y. T., and Snedeker, J. (2013). The use of lexical and referential cues in children's online interpretation of adjectives. *Dev. Psychol.* 49:1090. doi: 10.1037/a0029477

Hughes, M. E., and Allen, S. E. (2013). The effect of individual discourse-pragmatic features on referential choice in child English. *J. Pragmat.* 56, 15–30. doi: 10.1016/j.pragma.2013.05.005

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., and Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *J. Psycholinguist. Res.* 29, 597–626. doi: 10.1023/A:1026468209238

Jackendoff, R. (1978). "Grammar as evidence for conceptual structure," in *Linguistic Theory and Psychological Reality.* eds M. Halle, J. Bresnan, and G. Miller (Cambridge, MA: MIT Press), 201–228

Järvikivi, J., and Pyykkönen-Klauck, P. (2019). Changes in visual presence influence pronoun comprehension in children. Manuscript. doi: 10.13140/RG.2.2.17076.81286

Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., and Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: tracking eye movements to visually presented anaphoric referents. *Lang. Cogn. Neurosci.* 29, 877–892. doi: 10.1080/01690965.2013.804941

Järvikivi, J., Van Gompel, R. P., and Hyönä, J. (2017). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *J. Psycholinguist. Res.* 46, 525–550. doi: 10.1007/s10936-016-9451-1

Järvikivi, J., Van Gompel, R. P., Hyönä, J., and Bertram, R. (2005). Ambiguous pronoun resolution contrasting the first-mention and subject-preference accounts. *Psychol. Sci.* 16, 260–264. doi: 10.1111/j.0956-7976.2005.01525.x

Kaiser, E., and Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in finnish: evidence for a form-specific approach to reference resolution. *Lang. Cogn. Process,* 23, 709–748. doi: 10.1080/01690960701771220

Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition* 101, 1–42. doi: 10.1016/j.cognition.2005.08.002

Kehler, A., and Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theor. Linguist* 39, 1–37. doi: 10.1515/tl-2013-0001

Kidd, E., and Bavin, E. L. (2002). English-speaking children's comprehension of relative clauses: evidence for general-cognitive and language-specific constraints on development. *J. Psycholinguist. Res.* 31, 599–617. doi: 10.1023/A:1021265021141

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

Kidd, E., Stewart, A. J., and Serratrice, L. (2011). Children do not overcome lexical biases where adults do: the role of the referential scene in garden-path recovery. *J. Child Lang.* 38, 222–234. doi: 10.1017/S0305000909990316

King, J., and Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *J. Mem. Lang.* 30, 580–602. doi: 10.1016/0749-596X(91)90027-H

Knoeferle, P., and Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cogn. Sci.* 30, 481–529. doi: 10.1207/s15516709cog0000_65

Knoeferle, P., and Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: evidence from eye-movements. *J. Mem. Lang.* 57, 519–542. doi: 10.1016/j.jml.2007.01.003

Knoeferle, P., Habets, B., Crocker, M. W., and Münte, T. F. (2007). Visual scenes trigger immediate syntactic reanalysis: evidence from ERPs during situated spoken comprehension. *Cerebral Cortex.* 18, 789–795. doi: 10.1093/cercor/bhm121

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., and Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active english sentences: evidence from event-related potentials. *Brain Lang.* 100, 223–237. doi: 10.1016/j.bandl.2005.12.006

Kuperberg, G. R., Sitnikova, T., Caplan, D., and Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cogn. Brain Res.* 17, 117–129. doi: 10.1016/S0926-6410(03)00086-7

Kushner, D, Wilhite, T. L., and Rees, J. (1987). *The Brave Little Toaster [Motion picture]*. United States of America: Hyperion Pictures.

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Langton, S. R., Watt, R. J., and Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends Cogn. Sci.* 4, 50–59. doi: 10.1016/S1364-6613(99)01436-9

Lee, R., Chambers, C. G., Huettig, F., and Ganea, P. A. (2017). "Children's semantic and world knowledge overrides fictional information during anticipatory linguistic processing," in *Proceedings of the 39th Annual Meeting of the Cognitive Science Society (CogSci 2017)* eds G. Gunzelmann, A. Howes, T. Tenbrink, and E. Davelaar (Austin, TX: Cognitive Science Society). 730-735

Lenth, R. (2018). *Emmeans: Estimated Marginal Means, aka Least-Squares Means.* R package version 1.12. Available online at: https://CRAN.R-project.org/package=emmeans (accessed March 01, 2018).

Lowder, M. W., and Gordon, P. C. (2014). Effects of animacy and noun-phrase relatedness on the processing of complex sentences. *Mem. Cogn.* 42, 794–805. doi: 10.3758/s13421-013-0393-7

Lowder, M. W., and Gordon, P. C. (2015). Natural forces as agents: reconceptualizing the animate–inanimate distinction. *Cognition* 136, 85–90. doi: 10.1016/j.cognition.2014.11.021

Mak, W. M., Vonk, W., and Schriefers, H. (2002). The influence of animacy on relative clause processing. *J. Mem. Lang.* 47, 50–68. doi: 10.1006/jmla.2001. 2837

Matin, E., Shao, K. C., and Boff, K. R. (1993). Saccadic overhead: information-processing time with and without saccades. *Percept. Psychophys.* 53, 372–380. doi: 10.3758/BF03206780

Miltsakaki, E. (2002). Toward an aposynthesis of topic continuity and intrasentential anaphora. *Comput. Ling.* 28, 319–355. doi: 10.1162/089120102760276009

Montgomery, J. W., Evans, J. L., Gillam, R. B., Sergeev, A. V., and Finney, M. C. (2016). "Whatdunit?" Developmental changes in children's syntactically based sentence interpretation abilities and sensitivity to word order. *Appl. Psychol.* 37, 1281–1309. doi: 10.1017/S0142716415000570

Nappa, R., and Arnold, J. E. (2014). The road to understanding is paved with the speaker's intentions: cues to the speaker's attention and intentions affect pronoun comprehension. *Cogn. Psychol.* 70, 58–81. doi: 10.1016/j.cogpsych.2013. 12.003

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant Child Dev.* 10, 3–18. doi: 10.1002/icd.239

Nieuwland, M. S., Martin, A. E., and Carreiras, M. (2013). Event-related brain potential evidence for animacy processing asymmetries during sentence comprehension. *Brain Lang.* 126, 151–158. doi: 10.1016/j.bandl.201 3.04.005

Nieuwland, M. S., and Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *J. Cogn. Neurosci.* 18, 1098–1111. doi: 10.1162/jocn.2006.18.7.1098

Porretta, V., Kyröläinen, A.-J., Van Rij, J., and Järvikivi, J. (2016). *VWPre: Tools for Preprocessing Visual World Data. R Package Version 0.9.6.* Available online at: https://CRAN.R-project.org/package=VWPre (accessed September 13, 2017).

Porretta, V., Kyröläinen, A.-J., Van Rij, J., and Järvikivi, J. (2018). "Visual world paradigm data: From preprocessing to nonlinear time-course analysis," in *Intelligent Decision Technologies 2017: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies (KES-IDT 2017) – Part II Smart Innovation, Systems and Technologies.* eds I. Czarnowski, R. J. Howlett, and L. C. Jain (Cham: Springer International Publishing), 268–277. doi: 10.1007/978-3-319-59424-8_25

Pyykkönen, P., and Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Exp. Psychol.* 57:5. doi: 10.1027/1618-3169/a000002

Pyykkönen, P., Matthews, D., and Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: a visual world study of pronoun resolution. *Lang. Cogn. Process* 25:115129. doi: 10.1080/01690960902944014

Rose, R. L. (2005). *The relative contribution of syntactic and semantic prominence to the salience of discourse entities* (Unpublished Ph.D.). Northwestern University, Evanston, IL, United States.

Schumacher, P. B., Roberts, L., and Järvikivi, J. (2017). Agentivity drives real-time pronoun resolution: evidence from German er and der. *Lingua* 185, 25–41. doi: 10.1016/j.lingua.2016.07.004

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., and Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition* 71, 109–148. doi: 10.1016/S0010-0277(99)00025-6

Snedeker, J., and Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cogn. Psychol.* 49, 238–299. doi: 10.1016/j.cogpsych.2004.03.001

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

Song, H. J., and Fisher, C. (2005). Who's "she"? - Discourse prominence influences preschoolers' comprehension of pronouns. *J. Memory Lang.* 52, 29–57. doi: 10.1016/j.jml.2004.06.012

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., and Sedivy, J. C. (2002). Eye movements and spoken language comprehension: effects of visual context on syntactic ambiguity resolution. *Cogn. Psychol.* 45, 447–481. doi: 10.1016/S0010-0285(02)00503-0

Spivey-Knowlton, M., and Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition* 55, 227–267. doi: 10.1016/0010-0277(94)00647-4

Szewczyk, J. M., and Schriefers, H. (2011). Is animacy special?: ERP correlates of semantic violations and animacy violations in sentence processing. *Brain Res.* 1368, 208–221. doi: 10.1016/j.brainres.2010.10.070

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863

Thal, D. J., and Flores, M. (2001). Development of sentence interpretation strategies by typically developing and late-talking toddlers. *J. Child Lang.* 28, 173–193. doi: 10.1017/S0305000900004621

Traxler, M. J., Morris, R. K., and Seely, R. E. (2002). Processing subject and object relative clauses: evidence from eye movements. *J. Mem. Lang.* 47, 69–90. doi: 10.1006/jmla.2001.2836

Traxler, M. J., Williams, R. S., Blozis, S. A., and Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *J. Memory Lang.* 53, 204–224. doi: 10.1016/j.jml.2005.02.010

Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition* 73, 89–134. doi: 10.1016/S0010-0277(99)00032-3

Trueswell, J. C., Tanenhaus, M. K., and Garnsey, S. M. (1994). Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *J. Mem. Lang,* 33, 285–318. doi: 10.1006/jmla.1994.1014

Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *J. Exp. Psychol.* 19:528. doi: 10.1037/0278-7393.19.3.528

Van Dijk, T. A. (2014). *Discourse and Knowledge: A Sociocognitive Approach.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107775404

Van Gompel, R. P. G., and Järvikivi, J. (2016). "Processing sentence structure and its effect on reference processing," in *Visually Situated Language Comprehension,* eds P. Knoeferle, P. Pyykkönen-Klauck, and M. Crocker (Amsterdam; Philadelphia, PA: John Benjamins), 83–126.

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

Van Rij, J., Wieling, M., Baayen, R. H., and Van Rijn, H. (2015). *itsadug: Interpreting Time Series, Autocorrelated Data Using GAMMs. R package version 2.3.* Available online at: https://rdrr.io/cran/itsadug/ (accessed September 23, 2017).

Vogels, J., Krahmer, E., and Maes, A. (2013). When a stone tries to climb up a slope: the interplay between lexical and perceptual animacy in referential choices. *Front. Psychol.* 4:154. doi: 10.3389/fpsyg.2013.00154

Vogels, J., Maes, A. A., and Krahmer, E. J. (2014). Choosing referring expressions in Belgian and Netherlandic Dutch: effects of animacy. *Lingua* 145, 104–121. doi: 10.1016/j.lingua.2014.03.007

Weckerly, J., and Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology* 36, 559–570. doi: 10.1111/1469-8986.3650559

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *J. Exp. Child Psychol.* 99, 75–95. doi: 10.1016/j.jecp.2007.10.004

Weighall, A. R., and Altmann, G. T. (2011). The role of working memory and contextual constraints in children's processing of relative clauses. *J. Child Lang.* 38, 579–605. doi: 10.1017/S0305000910000267

Welder, A. N., and Graham, S. A. (2006). Infants' categorization of novel objects with more or less obvious features. *Cogn. Psychol.* 52, 57–91. doi: 10.1016/j.cogpsych.2005.05.003

Wiig, E. H., Secord, W. A., and Semel, E. (2004). *Clinical Evaluation of Language Fundamentals—Preschool, Second Edition (CELF Preschool-2).* Toronto: The Psychological Corporation/A Harcourt Assessment Company.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction With R.* Boca Raton: Chapman & Hall/CRC Press. doi: 10.1201/9781420010404

Wood, S. N. (2016). *mgcv: Mixed GAM Computation Vehicle With GCV/AIC/REMLsmoothness Estimation. R package version 1.8-23.* Available online at: https://CRAN.R-project.org/package=mgcv (accessed October 01, 2017).

Zhang, L., and 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), 2593–2598.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., and Smith, G. (2009). *Mixed Effects Models and Extensions in Ecology With R.* New York, NY: Springer. doi: 10.1007/978-0-387-87458-6