Person-based prominence guides incremental interpretation: Evidence from obviation in Ojibwe

Christopher Hammerly[a,*], Adrian Staub[b], Brian Dillon[c]

Accepted for publication in *Cognition*, April 4, 2022

[a]Department of Linguistics, University of British Columbia, Vancouver, BC, Canada, V6T 1Z4
[b]Department of Psychological and Brain Sciences, University of Massachusetts Amherst, Amherst, MA, USA, 01003
[c]Department of Linguistics, University of Massachusetts Amherst, Amherst, MA, USA, 01003
[*]Corrpesonding Author: chris.hammerly@ubc.ca

Abstract Distinctions related to person and animacy have long been known to impact both the grammar and incremental processing in a way that can be described through "prominence" scales. We put the generalizability of these scales to the test by examining the processing effects of a typologically uncommon distinction known as *obviation*, which is found in Ojibwe, an Indigenous language of North America. Obviation contrasts the single most discourse-salient animate third person (PROXIMATE) with other non-salient third persons (OBVIATIVE). Using a visual world paradigm, we show that obviation influences parsing and interpretation commitments under incremental ambiguity: Proximate nouns are assumed to be the agent of an action, while obviative nouns do not lead to strong incremental commitments. This result parallels previous findings in other languages with distinctions related to animacy and person, supporting a theory where the effect of prominence information in processing is the result of a common set of constraints derived from the alignment of scales related to person, syntactic position, and thematic role.

Keywords: field psycholinguistics | Algonquian languages | filler-gap dependencies | person-animacy hierarchy | obviation | relative clauses

1 Introduction

All human languages have grammatical systems that distinguish different types of entities and concepts. One fundamental contrast is between types of CONVERSATIONAL PARTICIPANTS, which permits reference to the author and addressee of a linguistic act, and a wide variety of other NON-PARTICIPANTS. Grammatical distinctions within these non-participants, or third persons, are subject to cross-linguistic variation, with common contrasts separating abstract versus concrete concepts, inanimate versus living things, plants versus animals, animals versus humans, and humans based on social status or conversational salience (Comrie, 1989; Corbett, 1991; Creamer, 1974; DeLancey, 1981; Zwicky, 1977). These PERSON categories have a pervasive influence on many key aspects of language including word order (Young and Morgan, 1987), agreement (Dawe-Sheppard and Hewson, 1990; Preminger, 2014), case (Dixon, 1994; Silverstein, 1976), and argument co-occurrence (Coon and Keine, 2020). They are also used by comprehenders to constrain real-time linguistic interpretation (Bornkessel-Schlesewsky and Schlesewsky, 2009; Foley, 2020; Hammerly, 2020).

Our goal in the present paper is to bring the literatures on the grammar and processing of person into closer alignment by investigating the real-time comprehension of OBVIATION, a typologically rare distinction between different types of third persons found in the Algonquian family, including Border Lakes Ojibwe (*Anishinaabemowin*), a Central Algonquian language spoken near the Rainy River in Northern Minnesota and Northwestern Ontario (Valentine, 1994). Obviation contrasts the single most prominent animate third person noun (PROXIMATE) and those third persons nouns that are peripheral to the discourse (OBVIATIVE) (Bloomfield, 1962). This marking is distinct from other discourse-oriented systems such as focus and topicalization (Dahlstrom, 2014) and is obligatory in contexts where more than one third person is mentioned (Aissen, 1997; Hockett, 1966).

Obviation occupies a middle ground between the marking of conversational participants and the more commonly observed non-participant distinctions such as those based on animacy. Like the author and addressee, the referent that occupies the proximate role is determined by the discourse context. This differs from a property such as animacy, where, generally speaking, it is the inherent lexical semantics of the noun that determines whether it is categorized as animate or inanimate. However, like animacy, contrasts in obviation specifically serve to carve the space of different types of non-participants. This unique mixture of properties makes the study of obviation of critical importance to understanding the nature of person systems in general, including how this information impacts incremental processing. As we describe below, animacy has well documented impacts on real-time interpretation in a wide range of languages. One possibility is that these effects are related to grammatical and typological models such as those based in the *person-based prominence scale* (Bornkessel-Schlesewsky and Schlesewsky, 2009). This view makes the prediction that the contextually-determined prominence conferred by the Algonquian obviation system will be used by comprehenders in similar ways to properties related to person and animacy.

We put this prediction to the test using visual world eye-tracking with first speakers of Ojibwe. Participants listened to sentences with a temporary ambiguity regarding the assignment of thematic roles, but no ambiguity about the obviation status of the nouns. During this critical period, we

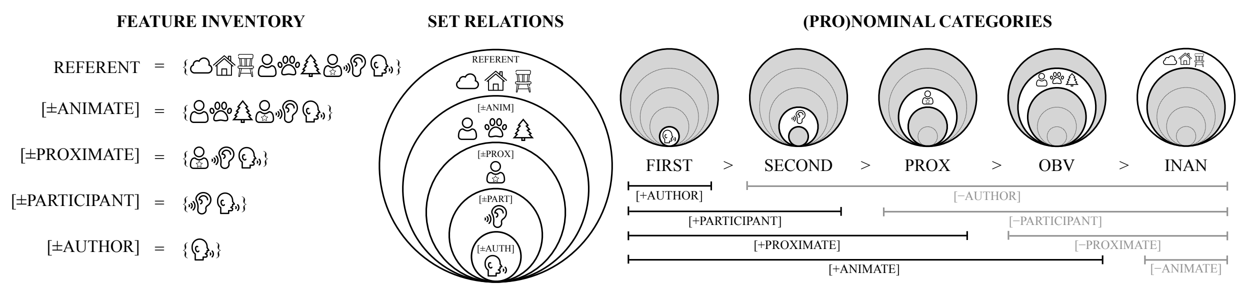

Figure 1: Schematization of the derivation of the five singular person categories of Ojibwe (FIRST, SECOND, PROXIMATE, OBVIATIVE, INANIMATE) from an inventory of four hierarchically related features ([±AUTHOR], [±PARTICIPANT], [±PROXIMATE], [±ANIMATE]). All icons are designed using resources from Flaticon.com

examined gaze patterns towards an array of images, which included two role-reversed depictions of the relationship between the nouns and the action being described (e.g. "Child greeting woman" versus "Woman greeting child"). We show that obviation displays parallel effects to those reported for animacy: Nouns assigned the more prominent proximate status show a pattern of looks that suggest speakers predictively assign them agent status, while the less prominent obviative nouns are not predictively assigned patient status. This result suggests that current accounts of how person and animacy influence real-time processing extend to obviation distinctions as well, suggesting a unified account of how personhood, animacy, and obviation are used in language comprehension.

1.1 Prominence and grammar

A key observation developed in the latter half of the 20th century is that the *features* that give rise to contrasts within the person reference space are hierarchically organized (Aissen, 2003; Dixon, 1994; Harbour, 2016; Silverstein, 1976; Zúñiga, 2006). This is schematized by the Venn diagram in Figure 1 for the distinctions present in Ojibwe singular nouns. It is this inventory of four "bivalent" features that allows for the derivation of the five *(pro)nominal categories* that refer to different sets of entities and concepts. Positively valued features carve out the space of possible referents with set intersection, finding a set that includes only those referents defined by the feature. Negatively valued features pick out the relative complement, finding a set that includes only those referents outside of the set defined by the feature. Ruling out contradictory combinations (e.g. [+AUTHOR] and [−PARTICIPANT]), this exhausts all logical feature-value combinations and derives all and only the five categories present in Ojibwe (Hammerly, 2020).

Beyond the accurate derivation of person categories, the subset relations between these four features form the basis of *person-based prominence scales* (Lockwood and Macaulay, 2012), the idea that the various categories related to person and animacy are ranked, as shown in (1).

(1) FIRST > SECOND > PROXIMATE > OBVIATIVE > INANIMATE

The ranking captures *implicational* relationships between the categories in how they interact with grammatical phenomena. For example, if a given category receives a particular case marking, all categories ranked above it will too, but those ranked below it may or may not (Dixon, 1994; Silverstein, 1976). This is derived by the fact that a given category necessarily shares at least one *positively* valued feature with those ranked higher that is *not* shared by those ranked lower. For example, FIRST, SECOND, and PROXIMATE share [+PROXIMATE], forming a natural class that excludes OBVIATIVE and INANIMATE.

This prominence scale has been important for describing grammatical systems across languages. In Ojibwe, the *direct-inverse voice* system has been described through the ranking of categories on the prominence scale (Dawe-Sheppard and Hewson, 1990; Macaulay, 2009). In transitive sentences with two animate third persons, one argument must be proximate and the other obviative—it is not possible for both to have the same obviation status. Overt singular proximate nouns appear in their bare form (e.g. *ikwe(w)* "woman (PROX)"), while overt obviative nouns are marked by a suffix *-an* (e.g. *ikwewan* "woman (OBV)"). Ojibwe is a radical *pro*-drop language, so both the subject and object are frequently null. This is the case in the sentences in (2). However, direct-inverse voice provides an unambiguous cue to which argument is proximate and which is obviative. In sentences where the higher ranked proximate person is acting on the lower ranked obviative person, a DIRECT marker appears, as with *-aa* in (2a). When this is reversed such that a lower ranked person is acting on a higher ranked person, the INVERSE marker appears, as with *-igoo* in (2b).

(2) a. o- waabam **-aa** -n
3- see -DIR -3′
'PROX sees OBV'

b. o- waabam **-igoo** -n
3- see -INV -3′
'PROX is seen by OBV'

The direct-inverse alternation occurs based not only on obviation, as illustrated above, but animacy and person as well. Furthermore, the direct-inverse alternation is not only a morphological marker that indicates the direction of an action—it is also associated with distinct syntactic configurations. As the translations in (2) indicate, the relationship between an argument's syntactic position and the thematic roles of agent (the "doer" of an action) and patient (the "undergoer" of an action) is similar to what is seen with the active-passive voice system of English. In direct sentences, the proximate *agent* is the grammatical subject. In inverse sentences, the proximate *patient* instead occupies the subject position (Bruening, 2005; Hammerly, 2020). This too can be described in terms of person-based prominence: It is always the higher-ranked proximate category that sits in the syntactically prominent subject position.

1.2 Prominence and incremental processing

A major task in language comprehension is to link nouns to syntactic positions and assign them thematic roles (Bornkessel-Schlesewsky and Schlesewsky, 2009; Ferreira, 2003; MacWhinney, Bates, and Kliegl, 1984). These linkages are particularly important in *filler-gap dependencies* such as those formed by relative clause (RC) structures, where a noun is not located in its canonical position. Consider the sentences in (3), where either the subject/agent (3a) or object/patient (3b) of the verb *quoted* has been displaced as the head of the RC.

(3) a. The reporter/report that __ quoted the senator
 b. The reporter/report that the senator quoted __

While it is generally the case in languages such as English that subject relatives clauses (SRCs) like those in (3a) are easier to comprehend than object relative clauses (ORCs) like those in (3b) (Holmes and O'Regan, 1981; Kwon, Lee, Gordon, Kluender, and Polinsky, 2010; Lau and Tanaka, 2021; Sasaki, Foley, Pizarro-Guevara, Silva-Robles, Toosarvandani, and Wagers, 2022; Wanner and Maratsos, 1978), numerous studies have found that this "subject advantage" only holds when animate nouns (e.g. *reporter*), but not inanimate nouns (e.g. *report*), are head of the RC (Gennari and MacDonald, 2008; Mak, Vonk, and Schriefers, 2002; Traxler, Williams, Blozis, and Morris, 2005; Wagers and Pendleton, 2016).

The subject advantage and its modulation by factors such as animacy is thought, at least in part, to be due to incremental expectations about how a sentence will unfold (Gennari and MacDonald, 2008, 2009; Sasaki et al., 2022; Wagers and Pendleton, 2016). When an animate noun is encountered, it is predicted to occupy the subject position and be assigned the agent role. If this prediction is foiled, as with an ORC continuation, the parse and interpretation must be revised, generating increased processing demands compared to an SRC continuation, which will match the predicted syntactic and thematic structure. This results in the observed subject advantage. In contrast, inanimate nouns do not appear to engender strong predictions in either direction, so both SRC or ORC continuations are on equal footing, leading to a lack of subject advantage in these cases. These effects are not limited to animacy: First and second person arguments can also alter the difficulty associated with ORCs and SRCs (Heider, Dery, and Roland, 2014; Reali and Christiansen, 2007). In particular, SRCs with a first or second person direct object (e.g. *the senator that saw me*) may be harder to process than ORCs with a first or second person subject (e.g. *the senator that I saw*). Furthermore, while not directly related to RC processing, previous studies with speakers of Odawa (Nishnaabemwin), an Eastern dialect of Ojibwe, have shown that patterns of production (Christianson and Ferreira, 2005) and offline comprehension (Christianson and Cho, 2009) are sensitive to animal versus human distinctions: While inverse sentences are less commonly produced and more difficult to understand than direct sentences when a human noun is the agent, this difference disappears when an animal noun is the agent.

What gives rise to person- and animacy-based expectations? The possibility that we explore here is that person-based prominence rankings are used directly in language comprehension (for an alternative view addressed in the Discussion, see the *Easy First* principle proposed by MacDonald (2013)). This claim is key to the Extended Argument Dependency model (Bornkessel-Schlesewsky and Schlesewsky, 2009), which proposes that the language comprehension system seeks to identify the most agent-like argument as quickly as possible. To do this, comprehenders use a set of cues that comprise features of prototypical agents, such as self-hood, animacy, definiteness, and even idiosyncratic lexical differences in order to incrementally evaluate the likelihood that a given argument is the actor within a clause (Alday, Schlesewsky, and Bornkessel-Schlesewsky, 2014; Bornkessel-Schlesewsky and Schlesewsky, 2009; Frenzel, Schlesewsky, and Bornkessel-Schlesewsky, 2015). As a result, this theory holds that the language comprehension system prioritizes interpretations where the person-based prominence scale is aligned with with grammatical and thematic role hierarchies (Gattei, Dickey, Wainselboim, and París, 2015a; Gattei, Tabullo, París, and Wainselboim, 2015b; Wagers and Pendleton, 2016), as well as other hierarchies related to case (e.g. Bornkessel-Schlesewsky and Schlesewsky, 2009; Foley, 2020). On this view, processing is facilitated when there is alignment between more prominent categories, roles, and positions, and inhibited otherwise (Wagers and Pendleton, 2016). Optimality Theory (Prince and Smolensky, 1993) provides one specific system of *harmonic alignment* to formalize these constraints (see Aissen, 1997, 1999). This can be conceived of either as a set of pressures present specifically during production, which then determine probability distributions used by a comprehender (MacDonald, 2013), or, perhaps more parsimoniously, as a single system that governs both production and comprehension (Momma and Phillips, 2018). In either case, higher ranked categories like animate, first, and second will be preferred as agents or subjects while lower ranked categories like inanimate will be preferred as patients or non-subjects.

Homing in on RC structures—the focus of the present study—these preferences are congruent with the observed subject advantage for animate RC head nouns, but alone they incorrectly predict an *object* advantage for inanimate head nouns. Instead, we see a neutralization of difficulty for SRCs versus ORCs in these cases (e.g. Gennari and MacDonald, 2008; Mak et al., 2002; Traxler et al., 2005; Wagers and Pendleton, 2016). This neutralization can be accounted for by a competing pressure: A general preference to form shorter dependencies over longer dependencies, in order to be as conservative as possible with limited memory resources (De Vincenzi, 1991; Futrell, Levy, and Gibson, 2020). This pressure will create a preference for SRCs over ORCs that is independent of the person-based prominence ranking of the head noun. In the case of an animate noun, the expectation that an SRC will be upcoming is even further strengthened, leading to a clear subject advantage. In the case of an inanimate noun, we have a clash of preferences: Based on animacy, we expect an ORC, but based on dependency length minimization, we expect an SRC. This results in neither expectation having a particular edge, leading to the observed neutralization of the subject advantage.

Turning to obviation, Ojibwe presents an opportunity to test the generalizability of the person-based prominence scale by allowing us to target a portion of the scale that is not distinguished in most languages. We hypothesize that higher-ranked proximate nouns should show a subject/agent advantage, while lower ranked obviative nouns should show a neutralization of the effect. Furthermore, if these preferences are a function of *predictive* processing, we should observe evidence of these commitments as a sentence unfolds. As mentioned above, while previous studies have reported experimental results with speakers of Odawa (Christianson and Cho, 2009; Christianson and Ferreira, 2005), neither included measures of incremental processing. We employ a variant of the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy, 1995), where Ojibwe speakers listened to relative clauses while looking at a series of images on a touch screen monitor (for a recent discussion of the importance of intuitive and game-like designs in field psycholinguistics, see Anand, Chung, and Wagers, 2020). By examining gaze patterns over the course of temporarily ambiguous sentences, we can determine what incremental commitments are made based on the obviation status of a noun. Based on our hypothesis, we expect proximate nouns to be immediately interpreted as agents, even before their role is disambiguated. Obviative nouns should either be interpreted as patients or to show no clear preference, given the different pressures that might guide Ojibwe speakers' interpretations in real-time processing

2 Methods

All methods were approved by the Institutional Review Board at the University of Massachusetts Amherst. Participants gave informed written consent before participating and were paid $40 USD. Anonymized data, analysis scripts, experimental scripts, and both audio and visual stimuli can be found here: https://osf.io/u3j4m ([dataset] Hammerly, Staub, and Dillon, 2022)

2.1 Participants

Participants (N = 19) were recruited with the help of community liaisons at Nigigoonsiminikaaning (n = 7) and Seine River (n = 12) First Nations. Both communities are Treaty 3 nations located within a short drive east of Fort Frances in Northwestern Ontario. There were 16 participants included in the final analysis. Two participants were excluded for declining to have video recordings taken. One participant was excluded for low performance on fillers. The mean age was 61.3 years, with the youngest being 42 and the oldest 80. All participants were deemed by the relevant community liaison to be fluent speakers of Ojibwe and self-reported their first language as Ojibwe, with most using the language on a daily basis in familial and community gatherings. All participants were fluent in English as a second language. The average age of first exposure to English was 6.3 years, with the range varying from 4 to 10 years old. Exposure to English was invariably tied to residential school, either directly or indirectly.

2.2 Materials

The experiment was comprised of a fully crossed 2x2 within-subjects design with factors HEAD (OBVIATIVE vs. PROXIMATE) and VOICE (DIRECT vs. INVERSE), as exemplified in (4). All sentences began with a preamble in Ojibwe that translates to "Choose the picture with...".

(4) a. PROXIMATE, DIRECT

... **gichi-aya'aa** gaa-baapi'**-aa**-d ininiw-an
elder.PROX REL-laugh-DIR-3 man-OBV
'...the elder who is laughing at the man'

b. PROXIMATE, INVERSE

... **gichi-aya'aa** gaa-baapi'**-igo**-d ininiw-an
elder.PROX REL-laugh-INV-3 man-OBV
'...the elder who is being laughed at by the man'

c. OBVIATIVE, DIRECT

... **gichi-aya'aa-n** gaa-baapi'**-aa**-d inini
elder-OBV REL-laugh-DIR-3 man.PROX
'...the elder who the man is laughing at'

d. OBVIATIVE, INVERSE

... **gichi-aya'aa-n** gaa-baapi'**-igo**-d inini
elder-OBV REL-laugh-INV-3 man.PROX
'...the elder who the man is being laughed at by'

We take advantage of the fact that RCs in Ojibwe are temporarily ambiguous between a reading where the head noun is the agent versus patient of the embedded verb: The relationship between a noun and its thematic role is determined by the combination of obviation and voice—there are no cues to these relationships based on other factors such as word order. This period of ambiguity spans from when the head noun is encoded until the point where voice morphology (*-aa/-igo*) is reached, i.e., during encoding of the relativizer *gaa-* and the lexical stem of the verb.

Each participant saw 32 experimental trials randomly interspersed with 16 filler trials (see Appendix B for experimental item transcriptions). The sentences were constructed through fieldwork with a first speaker community consultant (paid $40 USD per hour). This confirmed the felicity of the relative clause construction and generated a list of familiar nouns and verbs that could be used in the items. The end result was a set of 6 nouns and 24 transitive verbs from which the final sentences were constructed by the experimenter, counterbalancing the combination of the two nouns and the verb by item and condition. All items were then deemed by the community consultant to be culturally appropriate and comprehensible, and were translated and explained so they could be accurately depicted with images. Voice recordings were created with the community consultant using a Zoom H5 Handy Recorder with a Shure SM93 lavalier microphone. The raw .WAV files were cut into individual files by item and condition, edited for volume balancing, and in some cases digitally slowed to ensure the critical analysis regions were as lengthy as possible.

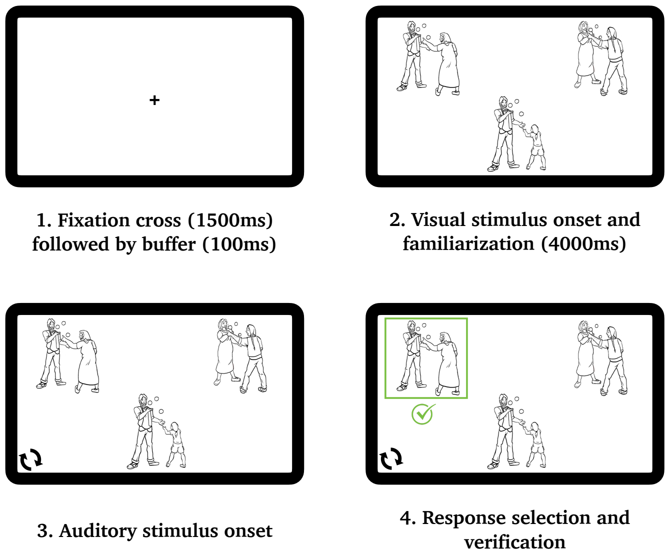

Figure 2: Schematization of experimental trials. In this example, the two role reversed images are on the top left and right, and the distractor is on the bottom.

An artist with connections to the Ojibwe community was contracted to create the visual stimuli, which comprised 150 drawings, 3 for each of the 48 trials, plus 6 for the practice session. All images were black and white line drawings depicting the transitive action in the sentence. For each experimental item, one image depicted the referent of the head noun as the agent (the *agent* image), one depicted the referent of the head noun as the patient (the *patient* image), and finally one entirely lacked a character corresponding to the head noun, but included the non-head noun character (the *distractor* image). For the distractor, whether the non-head noun character was depicted as an agent or patient was counterbalanced across items. One of the two role-reversed images ultimately corresponded to the correct response in the experimental trials, while the distractor image was always the correct response for the filler trials. Therefore each of the three types of images was the correct response on one-third of trials. Besides counterbalancing response options, this encouraged incremental interpretation, given that certain images could be ruled out prior to voice marking.

2.3 Procedure

The experiment was conducted at three different sites over a three week period in February 2020. Six participants were run at the cultural center at Nigigoonsiminikaaning and one at the round house. All Seine River participants were run in the home of the community liaison. Participants were run one at a time in a single session lasting approximately one hour. Participants were first given an overview of the experiment and the mechanics of the task. This was followed by a consent process and a brief demographic survey. The session continued with a practice session with 12 trials that introduced the six characters that appeared in the images to avoid confusion during the main task. This also served as an opportunity to further detail the mechanics and goals of the task and ensure the auditory stimuli were being presented at the appropriate volume. All instructions were given orally in English by the first author.

Participants then engaged in the main picture selection task, schematized in Figure 2, while their eye movements were recorded. The visual stimuli were presented on a 27” Acer T272HL Touch Screen Monitor and the auditory stimuli over USB-Powered Creative Pebble Desktop Speakers. Both visual and auditory presentation were controlled by a 2015 Macbook Pro running the PsychoPy2 experimental software. Preferential looking data were collected with a Logitech C920S HD Pro Webcam mounted at the top of the screen. The OBS Studio application was used to record the participant’s head and face as well as the experimental display, written into a single .MOV file at a rate of 30 FPS. Other behavioral data from the picture selection task were collected through PsychoPy2 and were output to a .CSV file. Interactions with the touch screen were time-locked to the onset of the auditory stimulus, with both the region and timing of every touch over the course of each trial being recorded.

Each trial was initiated by the participant, with a wait screen between each trial allowing breaks to be taken at any point (the screen read *touch the screen when ready to start* in Ojibwe; the only text that the participants saw in Ojibwe during the experimental session). The order of trials and distribution of fillers were randomized. The trial commenced when the participant touched the screen. Immediately following the registration of a touch, a fixation cross appeared at the center of the screen for 1500ms. Participants were instructed to train their gaze on this point. Following a 100ms blank screen buffer, the three visual stimuli appeared. One appeared in the bottom center of the screen, one in the upper left corner, and one in the upper right corner. The association between image type (agent, patient, and distractor) and location (left, right, bottom) was randomized on each trial, but recorded for use during data analysis. Participants were told to examine each of the images to familiarize themselves with the scenes and characters. The familiarization period lasted 4000ms, after which the sentence began to play.

Participants were instructed to select the image corresponding to the meaning of the sentence as soon as they could to encourage active parsing. Once selected, a green border would appear around the image, as well as a check mark either below the image (in the case of the two upper images) or above the image (for the lower image). At this point, participants could change their response by touching a different image, confirm their response by pressing the check-mark, or hear the sentence again by pressing a repeat button in the lower left corner of the screen. Participants were encouraged to go with their gut response and limit use of the repeat button. The images remained on the screen until the check mark was pressed, at which point the screen would go blank for 750ms and the wait screen would appear for the next trial. Following the final trial, the experimenter conducted a debriefing session to gather the participant’s impressions of the study. Participants were then paid and the session was concluded.

2.4 Picture selection analysis

The main analysis of the picture selection task focused on accuracy. Accuracy was measured using the final response on each trial. In order to fit a logistic mixed effects model for the analysis of the experimental trials, each trial was coded in a binary fashion as either correct or incorrect. For proximate-direct and obviative-inverse, the agent image was coded as a correct response. For proximate-inverse and obviative-direct, the patient image was coded as the correct response. All other responses (i.e. the role-reversed counterpart and the distractor) were coded as incorrect. While latency data was collected and is reported in Appendix A, there are a number of issues in interpreting this measure. First, there is not a single principled point when a response might be initiated. Participants could make a selection at any point from when the sentence began. Second, participants were able to change their answer, and also had to confirm their answer with a second button press, complicating the choice of how to define the end point of a response. Finally, significant variability in how motor responses were prepared when moving hands from their resting position to the screen strongly affects latency, making it difficult to interpret timing as measuring the main cognitive variables of interest. As a result, latency is presented without a statistical analysis and with an abundance of caution.

2.5 Preferential looking analysis

The .MOV file from each trial was hand-coded frame-by-frame to establish the gaze direction at each sampled time point. Coding was done by the experimenter using the VCode software, while blind to experimental condition and item number. Coding began at the onset of the fixation cross and continued until the first response touch was made. Each frame could be coded as (i) *right*, where the look was towards the image in the upper right corner; (ii) *left*, where the look was towards the image in the upper left corner; (iii) *bottom*, where the look was towards the image on the bottom; or (iv) *other*, which included looks towards the fixation cross at the start of the trial, blinks, saccades, off-screen looks, or otherwise uncategorizable frames. This direction-based coding was then transformed using the record of the experimental block so that looks were recategorized as towards the agent, patient, distractor, or other.

The main region of interest (ROI) is the *ambiguous region*, defined as the span of time from when the obviation status of the head noun becomes apparent up to when the direct/inverse marking occurs. In the proximate conditions, this started at the relativizer *gaa-*, where it becomes apparent that the head noun is unmarked. In the obviative conditions, this started at the onset of the obviative suffix on the noun. During this time, the thematic role of the head noun is unknown, but the obviation information is available. Two additional ROIs that are not integral to the core analysis were also calculated. One is the *resolution region*, which begins with the direct-inverse marking and ends whenever the first touch of the screen is made by the participant. This is where disambiguation occurs and the thematic roles can be established. The other is the *familiarization region*, which begins with the onset of the visual stimuli and ends with the preamble of the sentence. For each auditory stimulus, the critical time points were established by the experimenter by examining spectrograms in Praat. For the calculation of each region for each item, an additional 200ms was added to the end point on the assumption that there is a delay in saccade planning and execution such that looks during this additional time reflect earlier processing (e.g. Reichle, 2021, p. 372).

An issue for the analysis is that the length of each of the ROIs differs as a function of both item and condition. This is due to a number of factors including the experimental manipulation (e.g. the obviative suffix on the head noun makes the ambiguous region systematically longer in these conditions) and lexical variation (e.g. the embedded verb differs in length depending on the item). While this affects the *end point* of a region—that is, how long the region lasts—it is still possible to define a *relative starting point* that is shared across all trials for a given region. The relative starting point is treated as time zero for the analysis. For the critical ambiguous region, this is the point at which information about obviation is available. A semi-arbitrary cutoff was selected as the *maximal* end point. This was deemed necessary as the further from the zero point one gets, the smaller the data set to calculate the proportion of looks becomes, causing a decrease in the reliability of the measure. For the analysis of the ambiguous region, a cut was made at 1200ms, at which point 81% of trials are contributing (the shortest ambiguous region was 730ms). Finally, the data input to the statistical models included only those trials where the repeat button was *not* pressed, resulting in a loss of 11.3% of trials.

The analysis focuses on the ambiguous region. Given that there is no information about how the thematic dependencies resolve in this region, the analysis collapses across the contrast in voice. The goal of the analysis is to establish (i) *within* obviative and proximate conditions separately, whether there is a difference in the proportion of looks to the agent versus patient image (i.e. whether obviation is being used to predict the thematic role of the head noun); (ii) again *within* the proximate and obviative conditions separately, whether there is a difference in the proportion of looks between the agent versus distractor images and the patient versus distractor images (i.e. whether the distractor is being ruled out); and (iii) whether each of the above contrasts differ *between* the two head noun conditions (e.g. does the proportion of looks to the agent versus patient images differ as a function of obviation).

Times series data have a number of properties complicating statistical analysis, such as an inflation of Type 1 error due to multiple comparisons and the lack of independence (auto-correlation) of successive time points (see Huang and Snedeker, 2020; Stone, Lago, and Schad, 2020, for recent discussions). A solution for these issues can be found in non-parametric statistical tests. Following most closely the visual world analysis presented in Barr, Jackson, and Phillips (2014), we utilize a cluster-based permutation test. The test starts by identifying *clusters* of the sampled time points within the original data set that satisfy a chosen statistical criterion. Given the categorical nature of the gaze-based data, we used a generalized logistic mixed effects model to generate the test statistic z at each time point. Since a sample was taken every 33ms, we only considered clusters that span at least three frames, or 100ms. To construct the clusters, we chose a liberal cutoff of $p < .2$. The validity of the test against the null hypothesis distribution is not affected by this choice. For each of these clusters, a *cluster mass statistic* (CMS) is calculated by taking the sum of the individual test statistics within the cluster. For a given ROI, it is possible to identify multiple separate clusters.

Significance for each cluster was determined by a permutation test. We followed the synchronized permutation logic outlined by Barr et al. (2014). To this end, we randomized the labels of the relevant factor by subject. That is, for a given subject, either all labels of the relevant factor were reversed, or the original labels were retained. The idea behind the method is that because subjects may differ in how sensitive they are to the experimental manipulation, individual trials may not be exchangeable under the null hypothesis. Furthermore, the adopted permutation logic respects the original structure of the data, ensuring the validity of the null hypothesis distribution. With each of these randomized data sets, the *maximal* CMS is calculated and stored. Collectively, these maximal CMS's form a null hypothesis distribution, which can be used to determine a p-value for the clusters in the original data set. Under the conventional alpha threshold of $.05$, if fewer than 5% of the CMS values in the null hypothesis distribution are *more extreme than* the observed CMS, then the effect is deemed significant.

To isolate each relevant difference (agent versus patient, distractor versus agent, and distractor versus patient), the looking data were coded three different ways. For the difference between agent and patient looks, each time point was coded such that a look towards the agent image was 1, a look towards patient as 0, and distractor and other looks as NA. For the difference between distractor and agent, the distractor image was coded as 1, the agent 0, and the patient and other looks as NA. Finally, for the difference between distractor and patient looks, the distractor was coded as 1, the patient 0, and agent and other looks as NA.

To test the first two sets of questions (i.e., whether the proportion of looks between each of the three pairings of images differed *within* each head noun condition in the ambiguous region), six logistic regression models were specified, one for each of the three differences for the two levels of head noun obviation with subject and item random slopes and intercepts. These models were run at each time point in the original data set as well as the 2,000 alternative data sets under the cluster-based permutation procedure described above. This created six null hypothesis distributions for each of the six contrasts/models. In all cases, permutation proceeded with respect to the relevant image labels, with agent/patient, distractor/agent, or distractor/patient labels being randomly swapped, depending on the contrast at hand. To test the third question outlined above, that is, whether these three contrasts differed *between* levels of head noun obviation (i.e. to establish whether head noun condition modulates the proportion of looks to one image or the other), for each contrast in looks, a mixed effects logistic regression with a single fixed effect of head noun (Proximate = 1, Obviative = -1), along with maximal subject and item random effects, was used to generate a test statistic for the cluster-based permutation test. In this case, there were three models in total. The permutation procedure swapped the head noun condition labels to generate the alternative data sets.

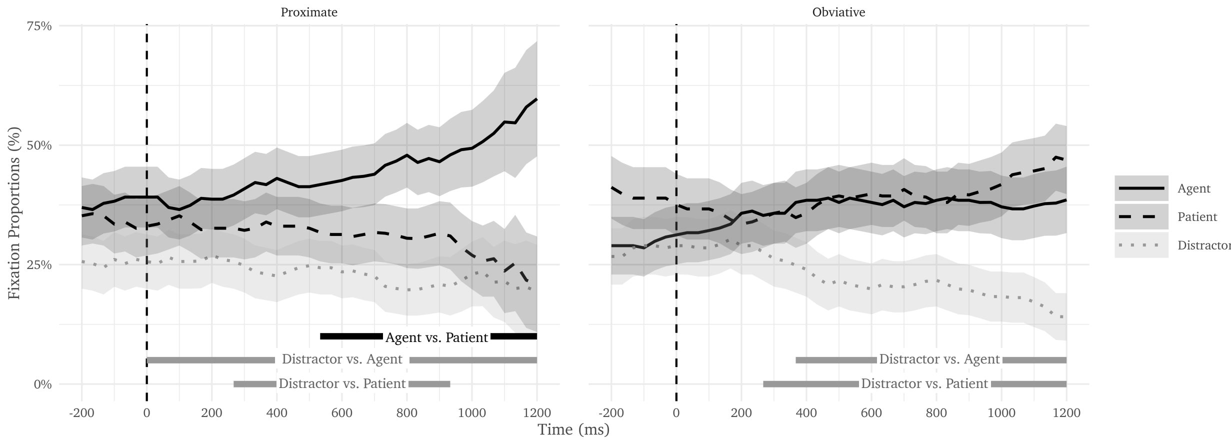

Figure 3: Average look proportions and 95% confidence intervals over time towards each of the three images in the ambiguous region (collapsed across levels of voice, which has not yet been encountered). Time zero marks the end of the head noun and the start of information related to obviation. The thick horizontal bars show significant clusters ($p < 0.05$) for contrasts in looks towards the agent versus patient, distractor versus agent, and distractor versus patient images.

3 Results

3.1 Preferential looking

We focus on the preferential looking results within the ambiguous region, shown in Figure 3. Results for the familiarization and resolution regions are given in Appendix A. The full results of the cluster-based permutation tests for the ambiguous region are given in Table 1. The main contrast of interest is between agent and patient looks. The tests revealed a significant cluster in the proximate head condition ($p = 0.013$) spanning the period of 533ms to the end of the region; during this period, there were more looks to the agent picture than the patient picture, when the head was proximate. No cluster, let alone a significant cluster, was found in the obviative head noun conditions. Furthermore, the test of the effect of head noun obviation on the proportion of agent looks was significant ($p = 0.005$) for a cluster spanning from 433ms to the end of the region, with more looks to the agent image, relative to the patient image, in the proximate head condition than in the obviative head condition. These results align with the view that obviation information is used to predictively assign thematic roles in a way that is consistent with the person-animacy hierarchy.

Furthermore, in both the proximate and obviative conditions, looks towards the distractor decreased over the course of the ambiguous region. For the distractor versus agent contrast, a significant cluster ($p = 0.001$) spanned the entire analysis region in the proximate condition, while in the obviative condition the cluster of significance ($p = 0.009$) ranged from 367ms to the end of the region. For the distractor versus patient looks, a significant cluster ($p = 0.010$) was found spanning 267–933ms for proximate heads and a significant cluster ($p < 0.001$) that spanned the time slice

Comparison	Head obviation	Cluster (ms)	CMS (z)	p-value
Agent v. Patient	Proximate	533–1200	48.54	*0.013
	Obviative	—	—	—
Distractor v. Agent	Proximate	0–1200	−112.39	*0.001
	Obviative	367–1200	−74.17	*0.009
Distractor v. Patient	Proximate	0–133	−7.60	0.185
		267–933	−38.52	*0.010
	Obviative	0–100	−5.96	0.221
		267–1200	−96.87	*<0.001

Effect of Head Obviation	Cluster (ms)	CMS (z)	p-value
Agent v. Patient	433–1200	55.55	*0.005
Distractor v. Agent	—	—	—
Distractor v. Patient	0–1200	29.02	0.078

Table 1: Results of cluster permutation tests. All located clusters are shown, along with the CMS with significance values determined from the constructed null hypothesis distribution.

Head	Voice	Agent	Patient	Distractor
Proximate	Direct	.891 (.025)	.039 (.019)	.070 (.022)
	Inverse	.273 (.052)	.664 (.052)	.063 (.016)
Obviative	Direct	.539 (.088)	.406 (.078)	.055 (.020)
	Inverse	.672 (.076)	.242 (.079)	.086 (.039)

Table 2: Mean response proportions and by-subjects SEM by condition for agent, patient, and distractor images.

starting at 267ms to the end of the region was found for obviative heads. This provides evidence that gaze data can provide a window into incremental interpretation in both conditions, since the distractor image, which lacks the character corresponding to the head noun, is being ruled out.

3.2 Picture selection

Box plots with by-participant accuracy in the picture selection task, broken down by the four experimental conditions, are given in Figure 4. The mean response proportions for each of the three image types, also split by condition, can be found in Table 2. Overall, relatively few of the responses were associated with the distractor image, indicating that participants had little trouble connecting the characters in the sentence to the images. Consistent with this, accuracy on the fillers, where the distractor image was the correct response, was 92%.

Effect	z	p-value
HEAD	3.39	*<0.001
VOICE	0.60	0.548
HEAD:VOICE	3.67	*<0.001

Table 3: Results of logistic regression on picture accuracy selection data.

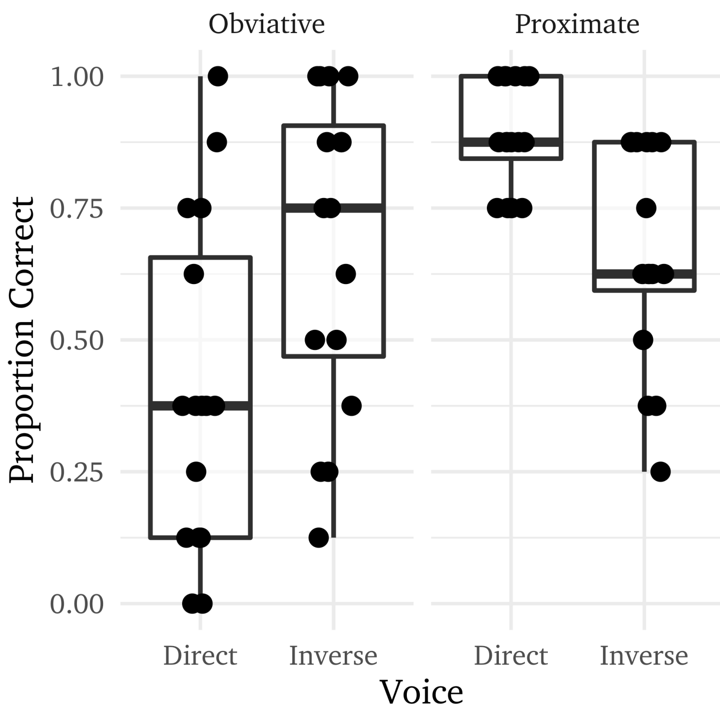

Figure 4: Box plots of by-participant response accuracy in the picture selection task, split by the four experimental conditions. Correct responses vary as a function of condition, with the agent image being correct with proximate/direct and obviative/inverse, and the patient image being correct with proximate/inverse and obviative/direct.

The full results of the logistic regression on response accuracy are given in Table 3. As suggested by Figure 4, the model revealed a main effect of head noun ($p < 0.001$) such that proximate conditions were more accurate than obviative conditions overall, and a significant interaction of head noun and voice ($p < 0.001$). *Post hoc* tests of the difference in accuracy between levels of voice were significant for both obviative heads ($t(15) = 3.01, p = 0.009$) and proximate heads ($t(15) = -4.42, p < 0.001$) such that the accuracy in the inverse condition was higher compared to the direct condition with an obviative head, and lower with inverse compared to direct with a proximate head. In other words, accuracy was higher when the head noun was ultimately resolved to be the agent of the embedded verb compared to the patient.

4 Discussion

We carried out a preferential looking task with 16 first speakers of Border Lakes Ojibwe. Our primary empirical goal was to understand how proximate and obviative nouns were interpreted in temporarily ambiguous relative clauses. The results of this study contribute three major findings. First, we found an increase in anticipatory looks towards images where the head noun was depicted as the agent when the head noun was proximate, but no preference when the head noun was obviative. This mirrors previous findings for animacy, where animate nouns have been found to show a subject advantage that is neutralized with inanimate nouns (Gennari and MacDonald, 2008; Mak et al., 2002; Traxler et al., 2005; Wagers and Pendleton, 2016). Second, we found that picture selection responses were more accurate when the head noun was proximate compared to obviative. Third, we found more accurate picture selection responses when the head turned out to be the agent rather than the patient, which appeared as a crossover interaction between obviation and voice.

We propose that all three results can be captured by the interactions between the following three independently motivated processing pressures:

(5) a. **Agent First Preference**: Assume the first noun is the thematic agent (Bever, 1970; Bornkessel-Schlesewsky and Schlesewsky, 2009; Ferreira, 2003).

b. **Subject Gap Preference**: Assume the first noun is the syntactic subject (Frazier, 1987; Gennari and MacDonald, 2008; Wagers, Borja, and Chung, 2018; Wagers and Pendleton, 2016).

c. **Prominence Alignment Preference**: Align higher ranked nouns with the subject position and agent role and lower-ranked nouns with non-subject and non-agent roles (Aissen, 1997, 1999).

These three pressures compete during online processing, with the parser updating its current parse in order to “maximize incremental well-formedness” (Wagers et al., 2018), where the relative well-formedness of possible competing parses is defined by the preferences encoded in each of the above pressures. Note that these pressures are not posited as unique to speakers of Ojibwe—they are general constraints on language processing derived from scales related to thematic role (AGENT

> PATIENT), syntactic position (SUBJECT > OBJECT), and person-based prominence (e.g. PROXIMATE > OBVIATIVE or ANIMATE > INANIMATE) that are hypothesized to be active in all languages (for a previous application of these types of constraints to explain Odawa language production, see Christianson, 2001a,b). Languages may differ in how heavily these constraints are weighted (Goldwater and Johnson, 2003) leading to variation in effects. For example, in the Mayan language Mam inanimate nouns cannot be subjects when the object is a human or animal (Minkoff, 2000), while inanimate subjects in the same context are simply dispreferred in English. A learner's particular linguistic experience over the course of acquisition determines the relative weight of these principles.

The presence of anticipatory looks towards the agent image in the proximate conditions, but no preference in the obviative conditions, supports a theory where the Agent First and Prominence Alignment pressures compete. Following the Prominence Alignment Preference, higher ranked categories like proximate nouns are assumed to be agents prior to encountering bottom-up evidence that confirms this analysis—a prediction that dovetails with the Agent First Preference. In contrast, based on the Prominence Alignment Preference, the lower-ranked obviative nouns are assumed to be patients. This clashes with the Agent First Preference, resulting in the appearance of a lack of preference in the obviative conditions. A potential alternative explanation is that our participants were predicting direct voice on the verb, which would allow them to infer that a proximate head noun is an agent. But this hypothesis would incorrectly predict a fully symmetric pattern of looks during the ambiguous region, with comprehenders confidently interpreting the obviative head noun as a patient before encountering the voice morphology. Therefore, this hypothesis is not consistent with the data seen here.

The finding of higher picture selection accuracy in the proximate conditions can be understood as the appearance of the subject advantage emerging from the Subject Gap Preference (Holmes and O'Regan, 1981; Kwon et al., 2010; Wanner and Maratsos, 1978) and Prominence Alignment Preference. While either proximate or obviative nouns can be thematic agents, only proximate nouns are promoted from their position of interpretation and thematic role assignment to the subject position (Bruening, 2005; Hammerly, 2020). This link to the subject position eases demands on memory resources by creating a shorter dependency (for a detailed formalization, see Hammerly (2020)). This relative ease in processing is then reflected in an overall increase in accuracy in the proximate conditions compared to the obviative conditions, where only a more resource-intensive non-subject relative clause is grammatically possible.

The final result—higher picture selection accuracy when the head noun is ultimately interpreted as the agent—is in line with previous experimental findings with speakers of the related language Odawa, which found a preference for putting the agent first in production (Christianson and Ferreira, 2005) as well as higher accuracy when interpreting basic declarative sentences where the first noun was ultimately interpreted as the agent (Christianson and Cho, 2009). As argued at length in Christianson and Ferreira (2005), the result is especially striking since Odawa shows significant word order flexibility—speakers are not boxed in by a rigid requirement to put the agent first, but such a preference emerges nonetheless. Both the results of the current study and those previously reported for Odawa can be understood as the generic Agent First Preference, where the first encountered noun is assumed to be the agent.

However, this raises a critical question: if the Agent First Preference is acting independently of the Prominence Alignment Preference, as suggested by the picture selection accuracy results, then why is this bias not also reflected in the preferential looking results in the obviative conditions? This could boil down to a difference in the timing and nature of the measures: While preferential looking reflects rapid and largely automatic online behavior arising from algorithmic parsing decisions, picture selection reflects offline decisions that are likely subject to interpretation heuristics and task effects (for recent work that elaborates on this split, see Meng and Bader, 2021; Paolazzi, Grillo, and Santi, 2021a; Paolazzi, Grillo, Alexiadou, and Santi, 2019; Paolazzi, Grillo, Cera, Karageorgou, Bullman, Chow, and Santi, 2021b). For example, participants might initially attempt a full syntactic parse. However, in the obviative conditions, the general difficulty of comprehending a non-subject relative clause results in uncertainty about the validity of the perceived input, which culminates by the time the sentence has completed. Encountering extreme difficulty like this is known to open the door to a more shallow set of "good enough" interpretation heuristics across a variety of languages and constructions (Ferreira and Patson, 2007; Zhou and Christianson, 2016; Zhou, Yao, and Christianson, 2018)—in this case, a heuristic to treat the first encountered NP as the agent (Bever, 1970; Ferreira, 2003). This can create a striking mismatch between the interpretation reported by participants and the actual meaning of a sentence like the one observed here. While this provides a plausible account of the online and offline data based on existing work, it remains speculative at present. More work is necessary to definitively establish the nature of the apparent contrast between our online and offline results.

Finally, one alternative school of thought directly links differences in processing difficulty to frequency of use: higher-ranked persons such as first, second, and animate third occur more frequently as subjects than as objects, making a subject parse more frequent and therefore more expected (MacDonald, Pearlmutter, and Seidenberg, 1994; Reali and Christiansen, 2007). While such corpus data are not currently available for Ojibwe, previous work examining patterns of production of regular declarative clauses has shown that proximate nouns are more commonly encoded as agents compared to obviative nouns, resulting in a higher frequency of direct voice marking compared to inverse (Christianson and Ferreira, 2005; Sullivan, 2016). This aligns with the results of the current study, where proximate nouns were predictively treated as agents prior to direct evidence from voice marking.

Critically, experience-based accounts must provide an explanation for why a *particular* set of patterns arises in the first place, as well as why the same patterns appear independently in many distinct linguistic communities. This is commonly achieved by appealing to general pressures that shape language production, such as an *Easy First* bias, where "conceptually salient" nouns, such as those referring to animate beings, are more easily retrieved from memory, and therefore produced earlier in an utterance (MacDonald, 2013). Given that, in many languages, the subject position and agent role are associated with earlier positions in a sentence, a strong association is formed between subject, agent, and conceptually salient nouns.

However, experience-based accounts using linear-temporal principles like the Easy First bias face myriad empirical challenges. First, numerous studies have shown that the link between the frequency of a particular structure and comprehension difficulty may be looser than expected if frequency is the sole determinant of difficulty (Heider et al., 2014; Wagers and Pendleton, 2016). Second, Easy First incorrectly predicts animacy effects in conjunction (e.g. a preference for *crabs and sand* over *sand and crabs*), where none are found (Bock and Warren, 1985). These effects appear to be limited to cases where structural hierarchy is at play. Third, languages such as Kaqchikel Maya show a consistent alignment between animacy and subjecthood (i.e. animate arguments are always subjects), while evidence from both traditional and experimental fieldwork shows that VOS word orders are more easily processed than SVO, despite putting the animate subject in a later linear position in the sentence (Kiyama, Tamaoka, Kim, and Koizumi, 2013; Koizumi, Yasugi, Tamaoka, Kiyama, Kim, Sian, and Mátzar, 2014). Similarly, the preferred word order in direct-voiced main clauses in Ojibwe is VOS, placing the less prominent obviative object before the more prominent proximate subject (Hammerly, 2021; Sullivan, 2016). That said, an adaptation of an experience-based account that directly invokes a version of the Prominence Alignment Preference in production could potentially be compatible with our results and the wider literature on person-based prominence in processing. More research is necessary to evaluate this possibility.

Overall, the results support the generalizability of the person-based prominence scale for understanding incremental processing in a diverse set of the world's languages. The scale provides the means to understand why information related to person, animacy, and obviation all show the same types of effects in the processing of syntactic and thematic structure. It can be boiled down to a single principle, where higher ranked categories are privileged as agents/subjects compared to lower ranked categories. These constraints then interact with other pressures, such as one that prefers the formation of shorter dependencies to conserve limited memory resources, and another generic agent-first heuristic, giving rise to a complex set of principles that guide incremental processing commitments in a diverse set of typologically distinct languages.

CRediT authorship contribution statement **Christopher Hammerly**: Conceptualization, Methodology, Software, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Funding Acquisition. **Adrian Staub**: Writing – Review & Editing, Supervision. **Brian Dillon**: Writing – Review & Editing, Supervision, Funding Acquisition.

Acknowledgments Chi-miigwech to Nancy Jones, Don Jones, and Andrew Johnson, who assisted in the creation of auditory stimuli and the recruitment of participants for this study. Thanks to Elijah Forbes for creating the visual stimuli used in this study. Kate Stone gave early pointers for the statistical analysis, Becky Tollan provided advice on how to code the preferential looking data and rightly convinced me to add a distractor image, Steven Foley gave feedback on a very early version of the experiment, and Rajesh Bhatt served as part of Hammerly's dissertation committee, from which this work stemmed. Thanks to Kiel Christianson and an anonymous reviewer for helpful feedback on this manuscript. Thanks also to audiences at the 34th Annual Conference on Human Sentence Processing, LSA 2021, the 52nd Algonquian Conference, the Move & Agree forum in 2021, the University of Toronto Scarborough, and the University of British Columbia for comments and feedback.

Funding This work was supported by two grants from the National Science Foundation: A Graduate Research Fellowship (1451512) to Hammerly and a Doctoral Dissertation Research Improvement Grant (1918244) to Dillon & Hammerly. This work was also supported by a Dissertation Fieldwork Grant from the Graduate School at UMass Amherst awarded to Hammerly. All views and claims are those of the authors.

Supplementary material Please find experimental scripts, analysis scripts, and anonymized data here: https://osf.io/u3j4m/

References

Aissen, Judith. 1997. On the syntax of obviation. *Language* 73:705–750.

Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. *Natural Language & Linguistic Theory* 17:673–711.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21:435–483.

Alday, Phillip M, Matthias Schlesewsky, and Ina Bornkessel-Schlesewsky. 2014. Towards a computational model of actor-based language comprehension. *Neuroinformatics* 12:143–179.

Anand, Pranav, Sandra Chung, and Matthew Wagers. 2020. Widening the net: Challenges for gathering linguistic data in the digital age. *Response to NSF SBE* .

Barr, Dale J, Laura Jackson, and Isobel Phillips. 2014. Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General* 143:404–413.

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In *Cognition and the development of language*, ed. John R. Hayes, 279–362. New York, NY: Wiley.

Bloomfield, Leonard. 1962. *The Menomini language*. Yale University Press.

Bock, J Kathryn, and Richard K Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21:47–67.

Bornkessel-Schlesewsky, Ina, and Matthias Schlesewsky. 2009. The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. *Language and Linguistics Compass* 3:19–58.

Bruening, Benjamin. 2005. The Algonquian inverse is syntactic: Binding in Passamaquoddy. *Manuscript, University of Delaware* .

Christianson, Kiel. 2001a. Optimality theory in language production: The choice between direct and inverse in Odawa. *Linguistica Atlantica* 23:93–126.

Christianson, Kiel. 2001b. An OT approach to variation in Odawa production data. In *University of british columbia working papers in linguistics: Proceedings of wscla 6*, 25–50. Vancouver: University of British Columbia Working Papers in Linguistics.

Christianson, Kiel, and Hee Youn Cho. 2009. Interpreting null pronouns (pro) in isolated sentences. *Lingua* 119:989–1008.

Christianson, Kiel, and Fernanda Ferreira. 2005. Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98:105–135.

Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.

Coon, Jessica, and Stefan Keine. 2020. Feature gluttony. *Linguistic Inquiry* Early Access.

Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.

Creamer, Mary Helen. 1974. Ranking in Navajo nouns. *Navajo Language Review* 1:29–38.

Dahlstrom, Amy. 2014. Obviation and information structure in meskwaki. In *Papers of the Forty-sixth Algonquian Conference*.

Dawe-Sheppard, Audrey, and John Hewson. 1990. Person and gender hierarchies in Micmac. *Journal of the Atlantic Provinces Linguistic Association* 12:1–12.

De Vincenzi, Marica. 1991. *Syntactic parsing strategies in Italian: The minimal chain principle*, volume 12. Springer Science & Business Media.

DeLancey, Scott. 1981. An interpretation of split ergativity and related patterns. *Language* 57:626–657.

Dixon, Robert M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Ferreira, Fernanda. 2003. The misinterpretation of noncanonical sentences. *Cognitive psychology* 47:164–203.

Ferreira, Fernanda, and Nikole D Patson. 2007. The 'good enough' approach to language comprehension. *Language and Linguistics Compass* 1:71–83.

Foley, Steven. 2020. Case, agreement, and sentence processing in Georgian. Doctoral Dissertation, University of California, Santa Cruz.

Frazier, Lyn. 1987. Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory* 5:519–559.

Frenzel, Sabine, Matthias Schlesewsky, and Ina Bornkessel-Schlesewsky. 2015. Two routes to actorhood: lexicalized potency to act and identification of the actor role. *Frontiers in psychology* 6:1.

Futrell, Richard, Roger P Levy, and Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language* 96:371–412.

Gattei, Carolina A, Michael W Dickey, Alejandro J Wainselboim, and Luis París. 2015a. The thematic hierarchy in sentence comprehension: A study on the interaction between verb class and word order in Spanish. *The Quarterly Journal of Experimental Psychology* 68:1981–2007.

Gattei, Carolina A, Ángel Tabullo, Luis París, and Alejandro J Wainselboim. 2015b. The role of prominence in Spanish sentence comprehension: An ERP study. *Brain and language* 150:22–35.

Gennari, Silvia P, and Maryellen C MacDonald. 2008. Semantic indeterminacy in object relative clauses. *Journal of memory and language* 58:161–187.

Gennari, Silvia P, and Maryellen C MacDonald. 2009. Linking production and comprehension processes: The case of relative clauses. *Cognition* 111:1–23.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*.

Hammerly, Christopher. 2020. Person-based prominence in Ojibwe. Doctoral Dissertation, University of Massachusetts Amherst.

Hammerly, Christopher. 2021. A verb-raising analysis of the Ojibwe VOS/VSO alternation: Lessons for feature copying and movement. Unpublished Manuscript, University of Minnesota, Twin Cities.

[dataset] Hammerly, Christopher, Adrian Staub, and Brian Dillon. 2022. Person-based prominence guides incremental processing: Evidence from obviation in Ojibwe. Retrieved from osf.io/u3j4m.

Harbour, Daniel. 2016. *Impossible persons*. Cambridge, MA: MIT Press.

Heider, Paul M, Jeruen E Dery, and Douglas Roland. 2014. The processing of *it* object relative clauses: Evidence against a fine-grained frequency account. *Journal of Memory and Language* 75:58–76.

Hockett, Charles F. 1966. What Algonquian is really like. *International Journal of American Linguistics* 32:59–73.

Holmes, Virginia M, and J Kevin O'Regan. 1981. Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior* 20:417–430.

Huang, Yujing, and Jesse Snedeker. 2020. Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition* 200:104251.

Kiyama, Sachiko, Katsuo Tamaoka, Jungho Kim, and Masatoshi Koizumi. 2013. Effect of animacy on word order processing in Kaqchikel Maya. *Open Journal of Modern Linguistics* 3:203–207.

Koizumi, Masatoshi, Yoshiho Yasugi, Katsuo Tamaoka, Sachiko Kiyama, Jungho Kim, Juan Esteban Ajsivinac Sian, and Lolmay Pedro Oscar García Mátzar. 2014. On the (non) universality of the preference for subject-object word order in sentence comprehension: a sentence-processing study in kaqchikel maya. *Language* 722–736.

Kwon, Nayoung, Yoonhyoung Lee, Peter C Gordon, Robert Kluender, and Maria Polinsky. 2010. Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of prenominal relative clauses in korean. *Language* 86:546–582.

Lau, Elaine, and Nozomi Tanaka. 2021. The subject advantage in relative clauses: A review. *Glossa: a journal of general linguistics* 6.

Lockwood, Hunter T, and Monica Macaulay. 2012. Prominence hierarchies. *Language and Linguistics Compass* 6:431–446.

Macaulay, Monica. 2009. On prominence hierarchies: Evidence from Algonquian. *Linguistic Typology* 13:357–389.

MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in psychology* 4:226.

MacDonald, Maryellen C, Neal J Pearlmutter, and Mark S Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101:676.

MacWhinney, Brian, Elizabeth Bates, and Reinhold Kliegl. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of verbal learning and verbal behavior* 23:127–150.

Mak, Willem M, Wietske Vonk, and Herbert Schriefers. 2002. The influence of animacy on relative clause processing. *Journal of memory and language* 47:50–68.

Meng, Michael, and Markus Bader. 2021. Does comprehension (sometimes) go wrong for noncanonical sentences? *Quarterly Journal of Experimental Psychology* 74:1–28.

Minkoff, Seth. 2000. Animacy hierarchies and sentence processing. *The syntax of verb initial languages* 201:212.

Momma, Shota, and Colin Phillips. 2018. The relationship between parsing and generation. *Annual Review of Linguistics* 4.

Paolazzi, Caterina L, Nino Grillo, and Andrea Santi. 2021a. The source of passive sentence difficulty: Task effects and predicate semantics, not argument order. In *Passives cross-linguistically*, 359–393. Brill.

Paolazzi, Caterina Laura, Nino Grillo, Artemis Alexiadou, and Andrea Santi. 2019. Passives are not hard to interpret but hard to remember: evidence from online and offline studies. *Language, Cognition and Neuroscience* 34:991–1015.

Paolazzi, Caterina Laura, Nino Grillo, Claudia Cera, Fani Karageorgou, Emily Bullman, Wing Yee Chow, and Andrea Santi. 2021b. Eyetracking while reading passives: an event structure account of difficulty. *Language, Cognition and Neuroscience* 1–19.

Preminger, Omer. 2014. *Agreement and its failures*. Cambridge, MA: MIT Press.

Prince, Alan, and Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Technical report, Rutgers University Center for Cognitive Science, Piscataway, NJ.

Reali, Florencia, and Morten H Christiansen. 2007. Processing of relative clauses is made easier by frequency of occurrence. *Journal of memory and language* 57:1–23.

Reichle, Erik D. 2021. *Computational models of reading: A handbook*. Oxford University Press.

Sasaki, Kelsey, Steven Foley, Jed Pizarro-Guevara, Fe Silva-Robles, Maziar Toosarvandani, and Matt Wagers. 2022. Evidence for a universal parsing principle in Santiago Laxopa Zapotec. Retrieved from osf.io/2wgd8.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In *Grammatical categories in Australian languages*, ed. Robert M. W. Dixon. Atlantic Highlands, NJ: Humanities Press.

Stone, Kate, Sol Lago, and Daniel J Schad. 2020. Divergence point analyses of visual world data: applications to bilingual research. *Bilingualism: Language and Cognition* 24:1–9.

Sullivan, Michael. 2016. Making statements in Ojibwe: A survey of word order in spontaneous sentences. In *Papers of the Forty-Fourth Algonquian Conference: Actes du Congrès des Algonquinistes*, 329. SUNY Press.

Tanenhaus, Michael K, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268:1632–1634.

Traxler, Matthew J, Rihana S Williams, Shelley A Blozis, and Robin K Morris. 2005. Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language* 53:204–224.

Valentine, Jerry Randolph. 1994. Ojibwe dialect relationships. Doctoral Dissertation, The University of Texas at Austin.

Wagers, Matthew W, Manuel F Borja, and Sandra Chung. 2018. Grammatical licensing and relative clause parsing in a flexible word-order language. *Cognition* 178:207–221.

Wagers, Matthew W, and Emily Pendleton. 2016. Structuring expectation: Licensing animacy in relative clause comprehension. In *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, 29–46.

Wanner, Eric, and Michael Maratsos. 1978. An ATN approach to comprehension. In *Linguistic theory and psychological reality*, ed. Morris Halle, Joan Bresnan, and George A. Miller, 119–161. Cambridge, MA: MIT Press.

Young, Robert W, and William Morgan. 1987. *The Navajo language: A grammar and colloquial dictionary, revisd ed.*. Albuquerque: University of New Mexico Press.

Zhou, Peiyun, and Kiel Christianson. 2016. I "hear" what you're "saying": Auditory perceptual simulation, reading speed, and reading comprehension. *Quarterly Journal of Experimental Psychology* 69:972–995.

Zhou, Peiyun, Yun Yao, and Kiel Christianson. 2018. When structure competes with semantics: reading Chinese relative clauses. *Collabra: Psychology* 4.

Zúñiga, Fernando. 2006. *Deixis and alignment: Inverse systems in indigenous languages of the Americas*. Amsterdam: John Benjamins.

Zwicky, Arnold M. 1977. Hierarchies of person. In *Proceedings of the Chicago Linguistic Society*, volume 13, 714–733. Chicago: Chicago Linguistics Society.

A Additional figures

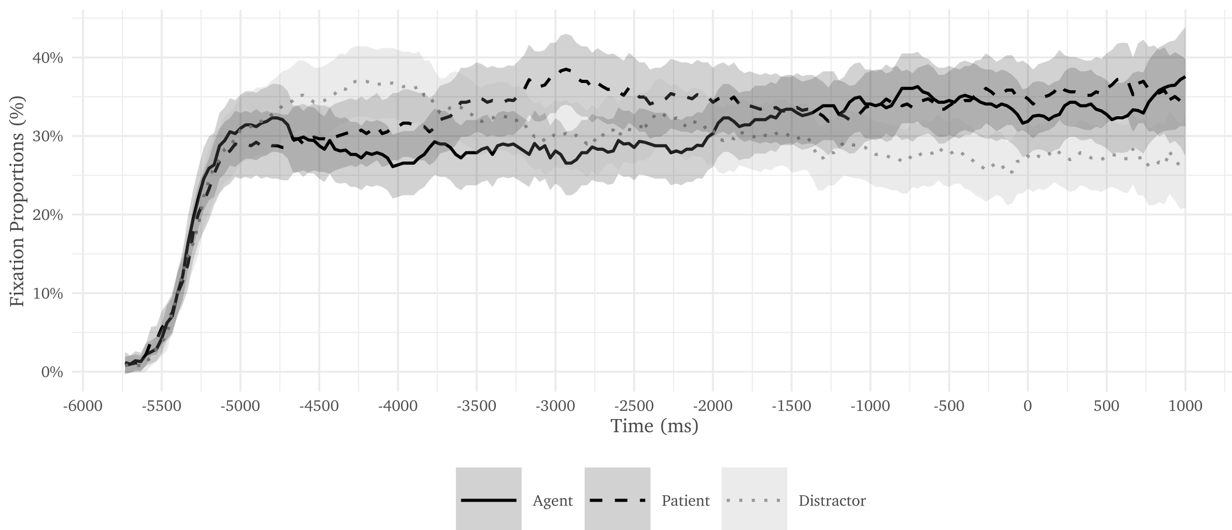

Figure A.1: Average look proportions over time in the familiarization region with 95% CIs. Time zero marks the onset of the audio stimulus.

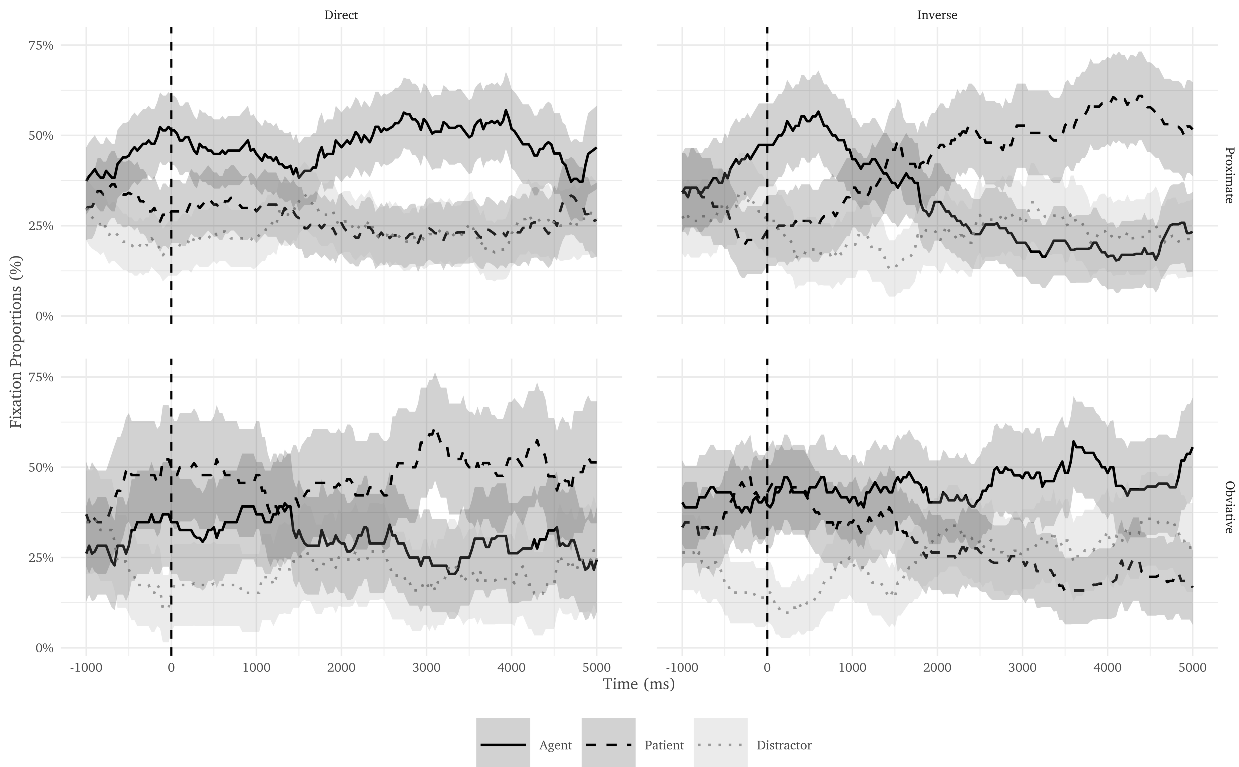

Figure A.2: Average look proportions over time in the resolution region with 95% CIs for correct responses only, split by head noun and voice. Time zero marks the onset of direct/inverse marking.

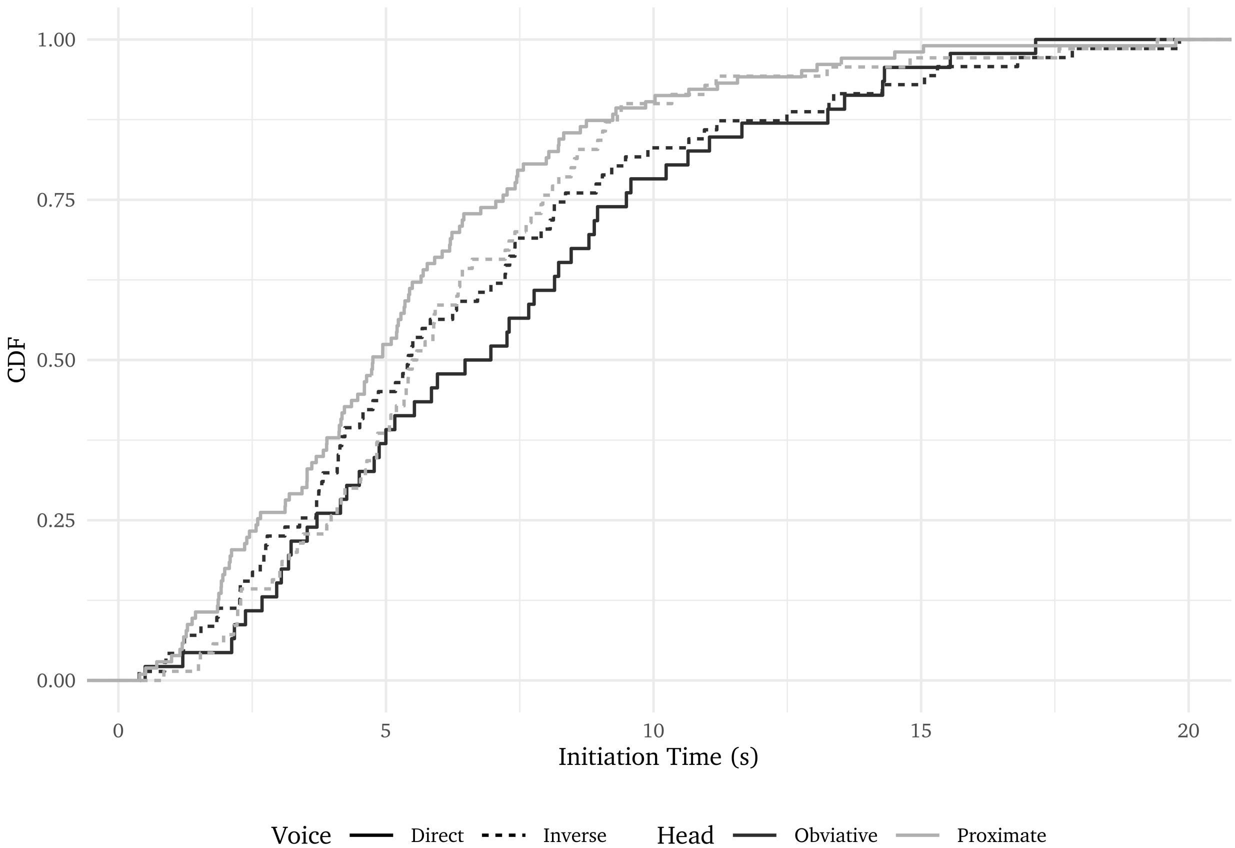

Figure A.3: Empirical Cumulative Density Function of response initiation time for correct responses made within 20s, time-locked to the end of the auditory stimulus. Includes trials where the first engagement with the touchscreen was ultimately the final response (i.e. does not include trials where the repeat button was pressed or the response was changed).

B Transcription of auditory stimuli

Item	Transcription	Translation
1	*ikwe gaa-binaakwe'waad abinoojiinyan*	'The woman who is combing the child'
2	*abinoojiinh gaa-ginagijiibinaad ininiwan*	'The child who is tickling the man'
3	*inini gaa-goshko'aad ikwewan*	'The man who is surprising the woman'
4	*oshki-inini gaa-nimikawaad oshkiniigikwewan*	'The teen boy who is dancing for the teen girl'
5	*gichi-aya'aa gaa-ojiimaad abinoojiinyan*	'The elder who is kissing the child'
6	*oshkiniigikwe gaa-anamikawaad gichi-aya'aan*	'The teen girl who is waving to the elder'
7	*ikwe gaa-jiisibinaad oshkiniigikwewan*	'The woman who is pinching the teen girl'
8	*abinoojiinh gaa-migwechiwi'aad oshkiniigikwewan*	'The child who is thanking the teen girl'
9	*inini gaa-binaakwe'waad oshkiniigikwewan*	'The man who is combing the teen girl'
10	*oshki-inini gaa-ginagijiibinaad ikwewan*	'The teen boy who is tickling the woman'
11	*gichi-aya'aa gaa-goshko'aad ininiwan*	'The elder who is surprising the man'
12	*oshkiniigikwe gaa-nimikawaad ikwewan*	'The teen girl who is dancing for the woman'
13	*ikwe gaa-ojiimaad gichi-aya'aan*	'The woman who is kissing the elder'
14	*abinoojiinh gaa-anamikawaad oshki-ininiwan*	'The child who is waving to the teen boy'
15	*inini gaa-jiisibinaad oshki-ininiwan*	'The man who is pinching the teen boy'
16	*oshki-inini gaa-migwechiwi'aad gichi-aya'aan*	'The teen boy who is thanking the elder'
17	*gichi-aya'aa gaa-amaji'aad abinoojiinyan*	'The elder who is waking up the child'
18	*oshkiniigikwe gaa-baapi'aad oshki-ininiwan*	'The teen girl who is laughing at the teen boy'
19	*ikwe gaa-mazinaakizwaad ininiwan*	'The woman who is photographing the man'
20	*abinoojiinh gaa-biminizha'waad ikwewan*	'The child who is following the woman'
21	*inini gaa-babaaminizhikawaad abinoojiinyan*	'The man who is chasing the child'
22	*oshki-inini gaa-izhinoo'waad ininiwan*	'The teen boy who is pointing at the man'
23	*oshkiniigikwe gaa-ashamaad abinoojiinyan*	'The teen girl who is feeding the child'
24	*gichi-aya'aa gaa-gikinoo'amawaad ikwewan*	'The elder who is teaching the woman'
25	*ikwe gaa-migoshkaaji'aad oshki-ininiwan*	'The woman who is pestering the teen boy'
26	*inini gaa-biindaakoonaad gichi-aya'aan*	'The man who is offering tobacco to the elder'
27	*abinoojiinh gaa-gagaanjinawe'aad oshki-ininiwan*	'The child who is irritating the teen boy'
28	*oshki-inini gaa-daanginaad gichi-aya'aan*	'The teen boy who is touching the elder'
29	*oshkiniigikwe gaa-zinigwaawiganebinaad ikwewan*	'The teen girl who is massaging the woman'
30	*gichi-aya'aa gaa-gikinjijiinaad oshkiniigikwewan*	'The elder who is hugging the teen girl'
31	*ikwe gaa-gaandinaad ininiwan*	'The woman who is pushing the man'
32	*abinoojiinh gaa-bakite'waad oshki-ininiwan*	'The child who is hitting the teen boy'

Table B.1: Example transcriptions of critical portion of experimental auditory stimuli. Table shows only proximate/direct conditions.