Journal of Memory and Language 102 (2018) 68–82

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Electrophysiological evidence for an independent effect of memory retrieval on referential processing

Hossein Karimi[a,*], Tamara Y. Swaab[b], Fernanda Ferreira[b]

[a] Pennsylvania State University
[b] University of California, Davis

ARTICLE INFO

Keywords:
Referential processing
Memory retrieval
Representational richness
Referential coherence
Nref

ABSTRACT

In this study, we show that the difficulty of re-activating and retrieving the representations of potential referents from memory (retrieval difficulty) influences referential processing, and that this effect is independent of the number of potential referents for a pronoun or the probability of possible referential interpretations (referential coherence). In two experiments, we varied retrieval difficulty by manipulating whether two referential candidates were modified by extra semantic information or not, creating representationally rich (modified) or bare (unmodified) referential candidates, respectively, and we measured event-related brain potentials (ERPs) on following pronouns. We observed a sustained frontal negative shift (the Nref effect) on pronouns following bare, and therefore difficult-to-retrieve, referential candidates relative to those following representationally rich candidates, regardless of the ambiguity of pronouns and the probability of either referential interpretation. Since referential coherence was held constant across the conditions, the results suggest that retrieval difficulty affects referential processing independently of coherence. We discuss the implications for memory-based theories of language processing.

Introduction

Pronouns such as *he* and *she* are commonly used in everyday language. Normally, pronouns carry little semantic information and derive their meaning from the entities they refer to, that is, their referents. Thus, processing a pronoun involves identifying a unique referent for it. For example, in *The actor had a long walk with the actress around the lake. He seemed a little nervous*, processing the pronoun *he* would in most cases result in co-indexation of *he* with the first-mentioned Noun Phrase (henceforth NP, i.e., *the actor*).

In this study, we draw a distinction between two potential sources of difficulty associated with resolving a referential dependency such as the one illustrated above. First, the difficulty of establishing a referential link could vary depending on the number of referential candidates that could potentially serve as the referent for a referring expression, and/or the probability of possible referential interpretations. Specifically, multiple or no potential referents could lead to more processing difficulty compared to a situation in which there is a single clear referent. Similarly, equi-probable referential interpretations could result in more processing difficulty than cases in which one interpretation is more likely than the other(s). We will henceforth call this property *referential coherence* (Nieuwland & Martin, 2017) because having one straightforward and unique referential interpretation results in a more coherent discourse compared to discourses in which multiple interpretations are possible or equally likely. The second source of difficulty associated with resolving a referential dependency might be based on how easily the memory representation associated with a referent can be retrieved from memory, as referential processing necessarily involves encoding a referential candidate in memory and then reactivating and retrieving that representation when a pronoun (or any other referring expression) is encountered (Dell, McKoon, & Ratcliff, 1983; Gernsbacher, 1989; Gerrig & McKoon, 1998; Lucas, Tanenhaus, & Carlson, 1990; MacDonald & MacWhinney, 1990; Sanford & Garrod, 1989, 2005). We will henceforth refer to this difficulty as *retrieval difficulty* as it captures the difficulty associated with retrieving the relevant representations from memory. The logic behind this distinction between referential coherence and retrieval difficulty is that even when a referring expression is coherent—that is, appears to have a single referent—the representation associated with that referent could be easy or difficult to retrieve depending on how activated it is in memory.

What happens when a referring expression has multiple rather than a single potential referent, and the discourse is consequently less coherent? Numerous studies have demonstrated that when multiple potential interpretations are plausible, the referential candidate that is relatively more activated in memory is retrieved faster and is taken to be the referent. For instance, it has been repeatedly shown that the

* Corresponding author. Department of Psychology, Pennsylvania State University, 140 Moore Building, University Park, PA, United States.
E-mail address: huk227@psu.edu (H. Karimi).

https://doi.org/10.1016/j.jml.2018.05.003
Received 1 July 2017; Received in revised form 4 May 2018
Available online 29 May 2018
0749-596X/

syntactic subject of a sentence is considerably more likely to be interpreted as co-referential with an ambiguous pronoun compared to the syntactic object (e.g., Arnold, 2001; Fletcher, 1984; Fukumura & Van Gompel, 2010; Givón, 1983; Gordon, Grosz, & Gilliom, 1993; Gundel, Hedberg, & Zacharski, 1993; Karimi & Ferreira, 2016a), presumably because the subject role confers more prominence on the associated NP (Gordon et al., 1993; Gundel et al., 1993), increasing its baseline activation level (e.g., Jäger, Engelmann, & Vasishth, 2017). The greater prominence of the syntactic subject is also consistent with the fact that it usually coincides with the "topic" of a sentence in English and topicality has also been shown to contribute to prominence (Ariel, 1990; Gundel et al., 1993; Rohde & Kehler, 2014). Similarly, ambiguous pronouns have been shown to be interpreted as referring to animate rather than inanimate NPs (Fukumura & Van Gompel, 2011) as well as to the semantically richer NPs (Karimi & Ferreira, 2016a). Since both animacy and semantic richness have been associated with greater activation levels in memory (animacy: Bock & Warren, 1985; Branigan, Pickering, & Tanaka, 2008; Rosenbach, 2008; semantic richness: Hofmeister, 2011; Troyer, Hofmeister, & Kutas, 2016), these results suggest that the referential candidate that is relatively more activated is taken to be the referent of an ambiguous pronoun. Thus, the relative activation levels of referential candidates directly influence discourse coherence by determining "what" (i.e., which referential candidate) is retrieved and taken to be the referent of the pronoun. In other words, when more than one referential interpretation is possible, the probability of a specific interpretation depends on the relative activation levels of the referential candidates, with the ultimate interpretation being consistent with the more easily retrievable candidate.

Numerous studies have investigated referential processing by recording event-related brain activity from the scalp (event-related potentials or ERPs). An effect of particular relevance to the present study is the *Nref* effect– a late and sustained frontal negativity that emerges in one condition relative to another and reflects referential processing difficulty[1] (e.g., Martin, Nieuwland, & Carreiras, 2012, 2014; Nieuwland, 2014; Nieuwland, Otten, & Van Berkum, 2007; Nieuwland & Van Berkum, 2006; Nieuwland & Van Berkum, 2008; Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003). The difficulty that has been observed in these studies may be attributable either to referential coherence, retrieval difficulty, or both. To test for retrieval difficulty it is necessary to hold constant all known factors that could contribute to referential coherence, and this is the major goal of the current study. In the remainder of this section, we first discuss the role of memory retrieval during referential processing, and then we will review relevant findings from previous studies in relation to our research question, arguing that current findings are equivocal with regards to the question whether ease of retrieval or referential coherence influence referential processing.

The role of memory retrieval during referential processing

Referential processing necessarily involves reactivating and retrieving the memory representations associated with the referential candidates from memory when a referring expression is encountered (Dell et al., 1983; Gernsbacher, 1989; Gerrig & McKoon, 1998; Lucas et al., 1990; MacDonald & MacWhinney, 1990; Sanford & Garrod, 1989, 2005). Therefore, one would expect ease of reactivation and retrieval from memory to influence referential processing and hence elicit an Nref effect (see Martin et al., 2012, 2014). This possibility is in line with cue-based retrieval theories of language processing. According to these theories, resolving a linguistic dependency such as the one between a pronoun and its referent involves a content-addressable memory search in which items needed for resolving the dependency are directly accessed via a cue-based retrieval mechanism (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Martin & McElree, 2009; McElree, 2006; McElree, Foraker, & Dyer, 2003; Van Dyke & McElree, 2006; also see Jäger et al., 2017).

Under these theories, a word whose processing depends on some previously encoded item is assumed to trigger the retrieval of that item for successful dependency formation. Crucially, the ease with which the target memory representation is retrieved is assumed to determine the difficulty of establishing the linguistic dependency (also see Gibson, 1998, 2000). In the original versions of cue-based retrieval theories, the retrieval of an item from memory is a function of its baseline activation level as well as the amount of interference that is experienced during its retrieval (e.g., Lewis & Vasishth, 2005; Lewis et al., 2006). Relevant to the current study, Jäger et al. (2017) performed a meta-analysis of current findings related to processing syntactic and referential dependencies and identified "competitor prominence" as an additional factor that influences the retrieval of the target item: The more prominent a competitor, the more difficult the retrieval of the intended (i.e., target) item. In the context of referential processing, cue-based retrieval predicts that the morpho-syntactic features of a pronoun such as its number and gender. (i.e., the retrieval cues) should activate the representations of all the referents that match those features to varying levels depending on the degree of match such that the representation that matches the most features should be activated the most and should eventually win the race for retrieval. Importantly, then, the ease with which the target representation (i.e., the referent) can be retrieved (determined by baseline activation, retrieval interference and competitor prominence) should affect referential processing difficulty and, therefore, the Nref.

Retrieval difficulty versus referential coherence

Previous studies have demonstrated that the Nref effect is elicited for less referentially coherent referring expressions relative to more coherent ones, suggesting that less coherent referring expressions are more difficult to process. However, these results are agnostic with respect to the distinction between retrieval difficulty and referential coherence. Broadly speaking, referential coherence has been manipulated via three factors in the current literature on referential processing: ambiguity, contextual bias in favor of possible referential interpretations, and feature mismatch between a pronoun and a single preceding referential candidate. Note that ambiguity and feature mismatch influence the number of potential referents for a pronoun, but contextual bias influences the probability of possible referential interpretations. Specifically, ambiguous referring expressions have been demonstrated to result in an Nref effect compared to unambiguous referring expressions, which is taken to reflect greater processing difficulty associated with referential ambiguity (e.g., Nieuwland & Van Berkum, 2006; Van Berkum et al., 1999 see also Nieuwland, Petersson, & Van Berkum, 2007; Nieuwland et al., 2007; Van Berkum et al., 2003).

Similarly, controlling for ambiguity, Nieuwland and Van Berkum (2006) manipulated verb bias and reported a greater Nref effect on ambiguous pronouns following contexts that provided weak support for either of two referential interpretations (e.g., *John hit David while he …*) compared to contexts that strongly supported one interpretation (e.g., *John hated David because he …*). In other words, referential processing was found to be easier when one referential candidate was more likely than the other to occur next given the preceding discourse. This is presumably because although the number of potential referents is formally the same, a strong bias makes one referential candidate much more probable as the referent than the other, leading to more coherent discourse than in contexts in which both referential interpretations are equally likely.

[1] The difficulty of referential processing has also been shown to result in N400 and P600 modulations (e.g., Swaab, Camblin, & Gordon, 2004; Ledoux, Gordon, Camblin, & Swaab, 2007; Camblin, Ledoux, Boudewyn, Gordon, & Swaab, 2007; Johns, Gordon, Long, & Swaab, 2014; Lamers, Jansma, Hammer, & Munte, 2006). However, since these studies were concerned with other aspects of referential processing such as the structural prominence of referents or the effects of referential failure, we will not discuss them here.

In addition, Nieuwland (2014) reported an Nref effect when a pronoun did not match a single preceding referential candidate in gender (*The boy thought that she would win the race*) relative to when it matched the gender of a referential candidate (*The boy thought that he would win the race,* but see Fiorentino, Covey, & Gabriele, 2018). Thus, when the number of potential referents within the current discourse representation was zero, referential processing was more difficult than when there was a unique referent for the pronoun.

Referential coherence can facilitate referential processing through at least two mechanisms: First, it is possible that less referential coherence leads to the need to entertain multiple referential interpretations, whereas greater referential coherence may result in the consideration of only a single interpretation (or at least the consideration of one interpretation more strongly than the other, Nieuwland & Van Berkum, 2006). For instance, unlike unambiguous pronouns, ambiguous pronouns are formally associated with multiple referents, which may lead to the entertaining of multiple referential interpretations. Similarly, weaker contextual bias in favor of either referential interpretation should make both interpretations more equally plausible, leading to stronger consideration of both. And finally, mismatching pronouns might result in consideration of out-of-context referential interpretations in addition to the infelicitous interpretation (Nieuwland, 2014).[2]

Second, less referential coherence might require additional inference for successful referential processing. For example, ambiguous pronouns need extra, context-based inference for pronoun resolution because the morpho-syntactic features of the pronoun are not sufficient to identify a unique referent, whereas unambiguous pronouns do not require such inference. Similarly, a weaker contextual bias necessarily requires more inference for referential processing than a stronger bias. And finally, mismatching pronouns might require inference to activate appropriate out-of-context referents whereas matching pronouns can be processed without such interference.

Crucially, however, all the three manipulations of referential coherence discussed above simultaneously result in variations in retrieval difficulty too. For instance, concerning ambiguity, because the retrieval cues of an ambiguous pronoun (such as gender and number) activates both referential candidates, the retrieval of either candidate is subject to interference during retrieval. However, such retrieval interference is greatly reduced when the pronoun is unambiguous because one referential candidate (e.g., the gender-matching one) is activated much more than the other. Similarly, with regards to the effect of contextual bias, when one referential candidate is more likely than the other to occur next, the more predictable candidate enjoys a greater activation level than the less predictable one and should therefore be easier to (re-) access (e.g., Levy, 2008). However, two equally predictable referential candidates have similar levels of activation in memory, rendering the retrieval of either more difficult. Another way to view this is that the competitor (non-target referential candidate) is necessarily more activated in weakly-biased than in strongly-biased contexts. Since higher activation of the competitor has been shown to complicate the retrieval of the target (Jäger et al., 2017), the contextual bias effect could be caused by retrieval difficulty. Finally, to the extent that mismatching pronouns result in co-referencing of the pronoun with out-of-context referents, the Nref effect elicited by mismatching pronouns could be due to retrieving a representation that has not been encoded during the processing of the current discourse, and is therefore less activated in memory and more difficult to retrieve.

[2] Note that in the case of gender-mismatching pronouns, the number feature of the pronoun (+ singular) still matches that of the preceding NP. Thus, based on cue-based retrieval theories, the infelicitous referent's representation should be activated (at least partially) despite the gender mismatch. As a result, in the case of gender mismatching pronouns, both the infelicitous referential interpretation as well as an out-of-context interpretation could be entertained.

The present study

We hypothesized that the ease with which the memory representations associated with referential candidates are retrieved from memory (i.e., retrieval difficulty) influences the difficulty of processing pronouns independent of the number or the probability of referential interpretations (i.e., referential coherence). We further hypothesized that retrieval difficulty alone would modulate the Nref. As discussed above, any effect of memory retrieval can be investigated only if the potential factors that could impact referential coherence are controlled for, namely, the number of potential referents (i.e., ambiguity and gender mismatch) as well as the probability of possible referential interpretations (i.e., contextual bias). In the current experiments, we kept both the number of potential referents and the probability of retrieval interpretations constant and manipulated only the absolute activation levels of referential candidates. We did this by varying the *representational richness* of both referential candidates– the amount of extra information attached to both referential candidates (see below).

Previous research has shown that semantically richer NPs are associated with stronger memory representations and are therefore easier to retrieve at a subsequent point. For instance, previous research has shown that semantically richer NPs are more likely to be interpreted as the referent of ambiguous pronouns (Karimi & Ferreira, 2016a), to be subsequently pronominalized (Karimi, Fukumura, Ferreira, & Pickering, 2014), and to occupy earlier sentence positions and/or more prominent syntactic roles during language production (Prat-Sala & Branigan, 2000; Yamashita & Chang, 2001). Semantically rich NPs have also been demonstrated to result in faster reading times when a subsequent constituent triggers their retrieval (Hofmeister, 2011; Hofmeister & Vasishth, 2014; Troyer et al., 2016). All these results have been interpreted as suggesting that richer representations are easier to retrieve from memory.

We examined the effect of representational richness on pronoun processing using sentences such as (1) and (2), in which an ambiguous pronoun is preceded by two viable referential candidates. In the Baseline condition, the two potential referents were bare determiner-noun sequences, and in the Representationally Rich condition, each referential candidate consisted of a determiner and a noun modified by a relative clause, increasing their representational richness, and therefore their ease of retrieval (e.g., Hofmeister, 2011; Hofmeister & Vasishth, 2014; Karimi & Ferreira, 2016a; Troyer et al., 2016).

(1) Sentence 1
  (a) **Baseline**: The actor walked away from the cameraman.
  (b) **Representationally Rich**: The actor who was visibly upset walked away from the cameraman who was critical of the show.
(2) Sentence 2
  After a while, he realized it was getting late and took a taxi home.

We measured ERPs to the critical pronoun in Sentence 2 as a function of the representational richness of the referential candidates in Sentence 1, focusing specifically on the Nref effect. Importantly, pronominal ambiguity, contextual bias, and feature (mis)match are kept constant across the two conditions (see, too, the results of our norming studies below). As a result, any Nref effect could only be attributed to the representational richness of the referential candidates, and by extension to retrieval difficulty alone. If referential candidates in the Representationally Rich condition are easier to retrieve from memory than those in the Baseline condition, and if this ease of retrieval influences the processing of the pronoun, then we would expect to observe the Nref effect on ambiguous pronouns following bare (unmodified) referential candidates relative to those following

representationally richer (modified) candidates.[3]

In contrast, if referential processing does not depend on ease of retrieval but only on the number and/or the probability of referential interpretations (i.e., referential coherence), there should be no reliable differences between the Baseline and the Representationally Rich conditions because the critical pronoun is equally coherent with the preceding discourse in both conditions (see the results of our norming studies below). In fact, because the linear distance between the critical pronoun and the referential candidates is necessarily longer when the referential candidates are modified by a relative clause than when they occur alone, we might actually observe an Nref effect on pronouns following modified relative to unmodified referential candidates. This possibility is consistent with previous research showing that linear distance complicates the retrieval of referential candidates presumably because the associated memory representations are more susceptible to decay when they are at a longer distance from the retrieval site (i.e., when the pronoun is encountered, Qiu, Swaab, Chen, & Wang, 2012; see also Gibson, 1998, 2000; Lewis & Vasishth, 2005; Lewis et al., 2006).

Experiment 1

The goal of this experiment was to establish whether the Nref is indeed sensitive to retrieval difficulty alone, independently of referential coherence, by comparing processing of pronouns following representationally rich (modified) vs. bare (unmodified) referential candidates. To isolate the potential effect of retrieval difficulty, we kept the pronoun ambiguous in both conditions, which controlled for the number of potential referents across conditions. Moreover, since both referential candidates are modified in the Representationally Rich condition (and not just one), the contextual bias, and therefore the probability of either referential interpretation, should remain the same across the two conditions. However, it is still possible that the semantic content of the added relative clauses might influence the activation of the referential candidates differently, changing which referential interpretation is adopted across conditions. For this reason, we conducted a norming study that tested to which referential candidate the ambiguous pronoun is taken to refer. Since ambiguous pronouns have been shown to be interpreted as referring to the NP that is relatively more activated and therefore easier to retrieve (among two viable NPs; e.g., Fletcher, 1984; Fukumura & Van Gompel, 2010; Karimi & Ferreira, 2016a), the norming study tested the *relative* activation levels of the two referential candidates by measuring the rate of resolving the ambiguous pronoun to each of the available NPs.

Norming study

Twenty-four undergraduate students from the participant pool of the University of California, Davis participated in the norming study in exchange for course credit. All the participants were native speakers of American English and reported no language-related disorders, and no exposure to any language(s) other than English before the age of five. We created 80 sentence pairs such as (1) and (2). Following previous studies (Koornneef & Van Berkum, 2006; Nieuwland & Van Berkum, 2006; Van Berkum, Koornneef, Otten, & Nieuwland, 2007), we presented the participants with discourse segments including sentences such as (1) plus all the words including and up to the critical pronoun in (2), resulting in discourse fragments such as (3). Participants were asked to provide a continuation for each discourse fragment and then explicitly indicate which referential candidate they talked about in their continuation.

(3)

(a) **Baseline**: The actor walked away from the cameraman. After a while, he …

(b) **Representationally Rich**: The actor who was visibly upset walked away from the cameraman who was critical of the show. After a while, he …

As is clear from (3), each critical sentence contained two referential candidates (e.g., *actor* and *cameraman*), one of which was always the syntactic subject and the first-mentioned noun phrase (NP1, *actor*), and the other was always the syntactic object and the second-mentioned noun phrase (NP2, *cameraman*). We also created 40 fillers which also contained two referential candidates but in the form of proper names (*Jack* and *Julie*). The two names had different genders in half of the fillers. The experimental sentences and the fillers were then distributed across 2 experimental lists so that each participant was exposed to only one version of each experimental item. The participants received the lists in Excel files on a desktop computer, and were asked to type their continuations next to each sentence (in a separate column), and then indicate which "character" (i.e., NP) they had "talked about" in their continuation (in a third column).[4] We recorded which referential candidate the participants indicated they had talked about in each continuation. We lost 17 (.008% of the) data points because some participants failed to identify a referent for some of the items.

Results of norming study

Table 1 reports the probability of interpreting the ambiguous pronoun as referring to each of the two potential referential candidates. As can be seen in this table, there was an overall tendency to interpret the pronoun as referring to the syntactic subject of the sentence (NP1) rather than to the syntactic object (NP2), which is consistent with numerous past studies (e.g., Arnold, 2001; Fletcher, 1984; Fukumura & Van Gompel, 2010; Givón, 1983; Gordon et al., 1993; Gundel et al., 1993; Karimi & Ferreira, 2016a). We conducted a repeated-measures ANOVA assessing the effect of Representational Richness on the choice between the two referential candidates (NP1 vs. NP2). The results revealed that the interpretation of the pronoun did not vary between the Baseline condition and the Representationally Rich condition (F (1, 23) = 1.53, p = .23). To test the potential effects of specific items, subjects and their interactions with our manipulation, we also ran a logit mixed-effects regression model with a full random effects structure (i.e., with random intercepts for both subjects and items, as well as by-subjects and by-items random slopes for Representational Richness; Barr, Levy, Scheepers, & Tily, 2013). Consistent with the ANOVA results, this regression model showed no effect of Representational Richness on the choice between NP1 and NP2 ($\beta = -.20$, $SE = .15$, $Z = -1.32$, $p = .18$).

Table 1
The percentages of choosing each referential candidate as the referent of the pronoun in each condition. Experiment 1.

Condition	NP1 Preference	NP2 Preference
Baseline	63.1%	36.9%
Representationally Rich	60.0%	40.0%

[3] Note that the verb phrases might also contribute to the richness of the first referential candidates. Moreover, the discourse markers preceding the pronoun (e.g., *After a while* in 2) differed in that some of them were temporal and some were causal in nature, which might differentially affect how the pronoun is resolved across items (see Kehler, Kertz, Rohde, & Elman, 2008). However, because the verbs and any discourse markers are identical across the two conditions, they cannot be responsible for any Nref effects that we might observe in one condition relative to the other in these experiments.

[4] In the original version of this paper, our norming studies for both experiments (see the norming study for Experiment 2 below) did not include the pronouns; participants received the first sentences (such as 1 above) and were asked to provide free continuations for them. The responses were then coded for which NP was talked about by naïve coders. However, two of the reviewers deemed those norming results as not representative of what was going on in the ERP experiments, and we therefore re-run both norming results including the critical pronouns in the prompt sentences.

Because this norming study indicates that the contextual bias does not differ across the two conditions, it appears that the probability of either referential interpretation is the same for each condition, suggesting that the *relative* activation levels of the two referential candidates is the same. Moreover, because the pronoun is always ambiguous, the number of potential referents is also the same across conditions, resulting in the same degree of referential coherence between the two conditions.[5] Consequently, both referential interpretations should be entertained to the same degree, and the same amount of inference should be applied across the two conditions. The only factor that varies between the conditions is the *absolute* activation levels such that *both* referential candidates should be more highly activated in the Representationally Rich than in the Baseline condition (e.g., Hofmeister, 2011; Karimi & Ferreira, 2016a; Karimi et al., 2014; Troyer et al., 2016).

Thus, if retrieval difficulty alone can influence referential processing, we should observe an Nref effect on pronouns following bare (unmodified) relative to representationally rich (modified) referential candidates. However, if referential coherence is necessary to produce the Nref effect, we should not observe a difference between the conditions.

ERP Experiment

Method

Participants. Forty-two undergraduate participants from the participant pool of University of California at Davis took part in the experiment in exchange for course credit (the participants were different people from those who took part in the norming study). They were all right-handed and were native speakers of American English and reported no neurological or language-related disorders, and no exposure to any other language(s) before the age of five. All participants signed a consent form prior to taking part in the experiment.

Stimuli. We used experimental stimuli such as (1) and (2). However, unlike in the norming study, the second sentence containing the critical pronoun was presented in full. The ambiguous pronoun never occurred as the first word of Sentence 2. Instead, a few discourse-connecting words preceded the pronoun so that the ERPs from the first sentence did not contaminate the critical Nref to the pronoun (see Appendix A for a sample of critical sentences for this experiment). The sentences were presented to the participants along with 224 other two-sentence discourse segments which served as fillers (the fillers were critical sentences for another independent study that did not investigate pronoun resolution). Half of the critical sentences and 60 of the fillers were tagged with a comprehension question in the form of a True/False statement. The questions for the filler sentences were about the general meaning of the sentence, while the questions for the critical sentences always assessed the resolution of the ambiguous pronoun. For example, the question following (1) and (2) was: *It was the actor who took a taxi home*.

The referential candidate mentioned in the questions was counterbalanced such that for half of the questions, the first referential candidate was mentioned (i.e., *the actor* in the example above), and for the other half the second referential candidate was mentioned (this would result in a question such as *It was the cameraman who took a taxi home*). This counterbalancing ensured that both *yes* and *no* responses were required for the same referential interpretation.

The 80 critical and 224 filler sentences were distributed across two experimental lists so that the two versions of each critical sentence did not occur within the same list. The items in each list were randomized for each participant, so that each participant read the stimuli in a different order. The experiment was programmed in Presentation software (Neurobehavioral Systems), which displayed the stimuli and sent out codes time-locked to the onset of critical pronouns in each condition.

Procedure. Participants were tested in an electrically-shielded, sound-attenuated booth. They were seated in front of a CRT computer screen placed 90 cm away from them. They were told that they would be presented with "two-sentence stories" and that their job was to read them for comprehension. The first sentence was presented on the screen all at once and the participants had unlimited time to read it. After they finished reading Sentence 1, they pressed the spacebar on the keyboard to trigger the presentation of Sentence 2, which was then automatically presented to them one word at a time at a rate of 500 milliseconds per word (each word stayed on the screen for 200 ms, followed by 300 ms of blank screen, producing an inter-stimulus interval of 300 ms). Sentence 2 was always preceded by a fixation cross to keep the participants' eyes fixated on the middle of the screen (where the words were to be displayed) and reduce eye-movement related artifacts. Once the presentation of Sentence 2 was completed, the comprehension question appeared on the screen (if there was one) and the participants were instructed to press the "m" button on the keyboard to indicate true and the "z" button to indicate false. They had unlimited time to answer the question. If there was no question tagged to the current sentence, a circle would appear in the middle of the screen and the participants could press either "m" or "z" to end the trial. As soon as "m" or "z" was pressed the next trial would start.

The participants were instructed to read the sentences carefully for comprehension and to do their best to answer the questions as accurately as possible. To minimize physiological artifacts on the EEG and consistent with standard practices in psycholinguistic research using ERPs, the participants were instructed to sit as still as possible and refrain from blinking and moving their eyes during the word-by-word presentation part of each trial. They were told they could blink during the first sentence and after the second sentence was over (which included the time when they answered the comprehension questions). The experimental session was not divided into separate blocks, but the participants were told they could stop to take a break after Sentence 2 was over when they felt they needed one. A practice session preceded the main experiment and consisted of six trials (three of which included comprehension questions), which allowed participants to become accustomed to the experiment. The experiment took approximately 2 hours to complete.

EEG recording. We recorded the electroencephalogram (EEG) from 29 tin electrodes mounted in an elastic cap (Electro-Cap International; Eaton, OH). Additional electrodes were attached below and to the side of each eye to record blinks and horizontal eye movements for later artifact rejection. All electrode impedances were kept below 5 kΩ. The EEG signal was amplified using a Synamps Model 8050 Amplifier (Compumedics Neuroscan) with a bandpass of .05–30 Hz. The EEG signal was digitally recorded at a sampling rate of 250 Hz.

Recorded EEG data from each participant were submitted to an independent component analysis (ICA) to isolate and remove EEG components associated with eye blinks. Single-trial waveforms were then screened for other artifacts such as amplifier drift, muscle and eye movements, and any epochs containing these artifacts were rejected prior to analysis (approximately 3% of all trials). Then, the data were re-referenced to the average of the left and right mastoids off-line and ERPs were calculated by averaging individual EEG epochs, time-locked to the presentation of the ambiguous pronoun in both conditions. All ERP epochs were 1500 ms long, extending from 300 ms prior to the presentation of the critical pronoun (baseline) to 1200 ms after the pronoun (epochs were baseline-corrected). ERP trials were sorted based on condition (Baseline vs. Representationally Rich) prior to off-line averaging.

[5] Note that the animacy, the information structure, and the relative semantic richness of the two NPs (within each condition) are also the same. Moreover, there would appear to be no motivation for positing out-of-context referential interpretations in our stimuli.

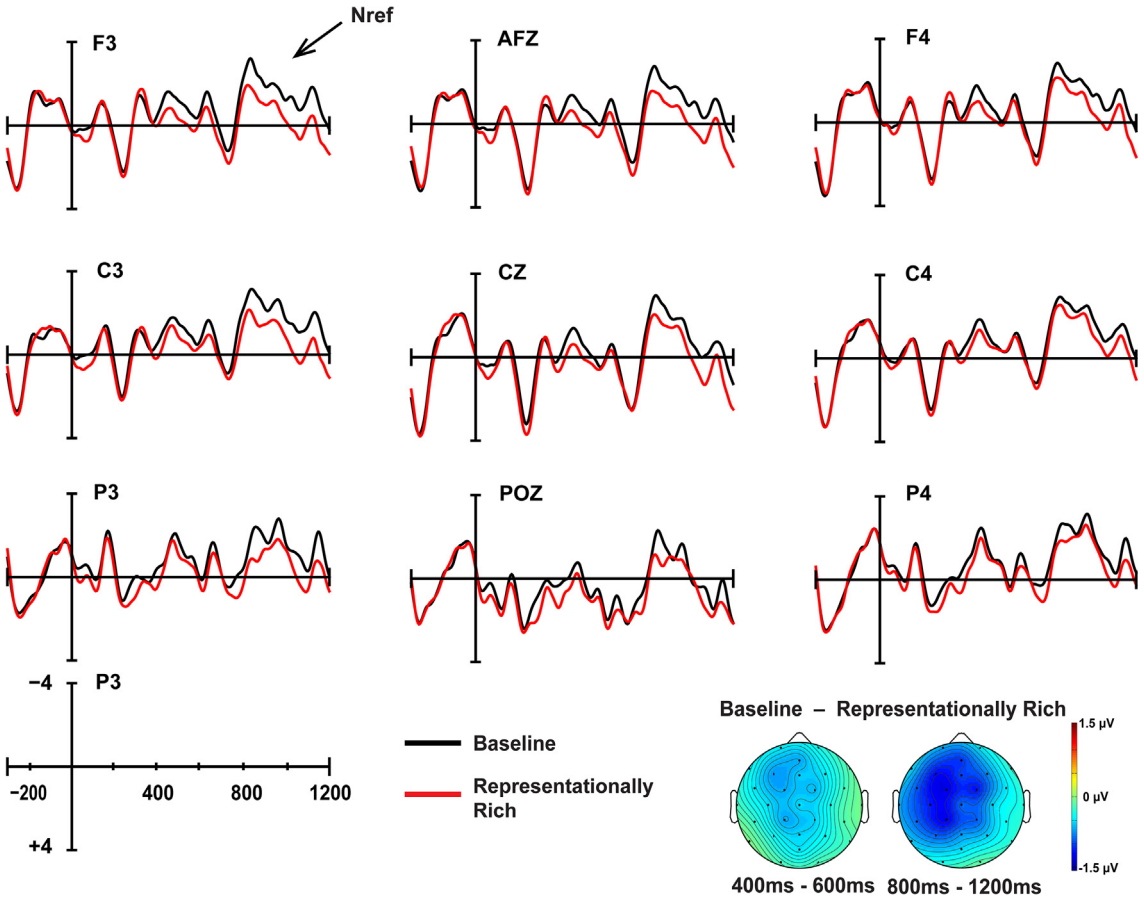

Fig. 1. Grand average brain waveforms time-locked to pronoun onset in the Baseline and the Representationally Rich conditions. Experiment 1.

Results. Fig. 1 shows the grand average brain waveforms for the Baseline and Representationally Rich conditions time-locked to the onset of the ambiguous pronoun in Sentence 2.

As can be seen in Fig. 1, we observed a frontal and sustained negative shift in the Baseline compared to the Representationally Rich condition emerging at about 400 ms after pronoun onset. The difference between the two conditions seemed to be maximal between 400 ms-600 ms and 800 ms-1200 ms following the onset of the pronoun. Based on this visual inspection and previous literature (e.g., Nieuwland, 2014; Nieuwland & Van Berkum, 2006; Nieuwland et al., 2007; Van Berkum et al., 2003), we performed our statistical analyses in these two time-windows.

We defined three brain regions to capture the topographical distribution of the effect. The "frontal" region included the following 12 electrodes: FP1, FP2, F3, F4, F7, F8, FC1, FC2, FC5, FC6, AFZ and FZ. The "central" region included the following 5 electrodes: C3, C4, T3, T4, CZ, and the "posterior" region included the following 12 electrodes: CP1, CP2, CP5, CP6, P3, P4, T5, T6, O1, O2, PZ and POZ. We then performed three separate ANOVAs testing the effect of representational richness within each brain region. Since the Nref has been shown to be a primarily frontal effect, we expected to find an Nref effect at the frontal regions. Table 2 reports the results of our statistical analyses.

As can be seen in Table 2, the main effect of Representational Richness was significant in the 400 ms-600 ms window in the frontal regions and marginally significant in the central regions, but not in the posterior region. In the 800 ms-1200 ms window, the effect was reliable in all three brain regions.

Unfortunately, we were unable to perform a standard analysis of the question-answering data due to a programming error which resulted in the loss of all item codes. This error was disappointing but not devastating, as our results are directly comparable to published results

Table 2
ANOVA results for the Representational Richness effect in the critical time windows, and within the three brain regions of interest. Experiment 1.

Time Window	Brain Area	Results
400 ms-600 ms	Frontal	$F(1,41) = 4.52$, $p = .04$
	Central	$F(1,41) = 3.87$, $p = .06$
	Posterior	$F(1,41) = 2.26$, $p = .14$
800 ms-1200 ms	Frontal	$F(1,41) = 8.82$, $p = .005$
	Central	$F(1,41) = 12.60$, $p = .001$
	Posterior	$F(1,41) = 5.86$, $p = .02$

because most previous studies on the Nref did not include any comprehension questions as part of the experimental protocol (e.g., Nieuwland & Van Berkum, 2006; Van Berkum et al., 1999, 2003). Nonetheless, for the second experiment we did include comprehension questions to assess the resolution of the pronoun, to allow us to examine the participants' final interpretations (see below).

Discussion. The results of Experiment 1 revealed that the ambiguous pronoun following bare (unmodified) referential candidates elicited an Nref effect at the fronto-central brain regions relative to ambiguous pronouns following representationally richer (modified) referential candidates, suggesting that the ease with which the representations associated with the referential candidates could be retrieved from memory (retrieval difficulty) influences referential processing independently of the number and the probability of potential referential interpretations (referential coherence). The pronoun was ambiguous in both the Representationally Rich as well as in the Baseline condition, and therefore the number of potential referents was the same across the conditions. In addition, based on the results of

the norming study, the contextual bias, and therefore the probability of each referential interpretation, was also the same across the two conditions. Consequently, the results are unlikely to have been caused by variations in referential coherence of the pronoun.

The factor that did vary across the two conditions was the *absolute* activation levels of the two referential candidates by virtue of the relative clauses added to the referential candidates. Thus, no matter which referential candidate was retrieved when the ambiguous pronoun was being processed, the retrieval process should have always been easier in the Representationally Rich relative to the Baseline condition. Note that unlike differences in *relative* activation levels, a difference in *absolute* activation levels cannot contribute to referential coherence because it does not affect the degree to which both referential interpretations are entertained across the two conditions, or to the amount of inference needed to arrive at a referential interpretation. In both the Baseline and the Representationally Rich conditions, the probability of interpreting the pronoun as referring to NP1 and NP2 was the same (as confirmed by our norming results). As a result, if both referential interpretations were entertained, they must have been entertained to the same degree across the two conditions. Similarly, the amount of inference needed to resolve the ambiguous pronoun would also have been the same across conditions.

An interesting aspect of our data that merits discussion is the fact that the referential candidates and the critical pronouns were farther apart in the Representationally Rich condition than in the Baseline condition, because they were separated by the post-nominal relative clauses. Nonetheless, the results showed easier referential processing for richer candidates. Thus, our results also suggest that representational richness overrides the effect of linear and temporal distance (and by extension memory decay), which is consistent with previous findings (Karimi et al., 2014; Karimi & Ferreira, 2016a; see Lewis & Vasishth, 2005; Lewis et al., 2006; and Gibson, 1998; Gibson, 2000 for the interactions between memory decay and retrieval).

However, it is possible to suggest an alternative explanation for the results of Experiment 1: Because there was generally more information to process when the referential candidates were both modified than when the NPs were bare, participants might not have committed to a referential interpretation in the Representationally Rich condition at all, and might have left the referential dependency unspecified. This could have been caused by a tendency to lower the burden imposed on processing resources by the greater syntactic and semantic complexity in this condition. Such a scenario would be in line with the Good Enough approach to language processing according to which linguistic dependencies are not always fully resolved (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007; Karimi & Ferreira, 2016b; Sanford & Sturt, 2002; Swets, Desmet, Clifton, & Ferreira, 2008; also see Christianson, 2016 for a review on good-enough language processing). For example, Swets, et al. (2008) showed that ambiguous syntactic dependencies where a relative clause could attach to either of two preceding NPs might sometimes remain unresolved. We conducted a second experiment to test this alternative explanation.

Experiment 2

To determine whether the results of Experiment 1 were due to underspecification of the referential dependency in the Representationally Rich condition, we added pronominal ambiguity as a second predictor, producing a 2×2 design: Representational Richness (Baseline vs. Representationally Rich) × Ambiguity (Ambiguous vs. Unambiguous). An example of a critical item is given in (4) and (5).

(4) Sentence 1
- (a) **Baseline_Ambiguous**: The actor walked away from the cameraman.
- (b) **Baseline_Unambiguous**: The actor walked away from the actress.
- (c) **Representationally Rich_Ambiguous**: The actor who was visibly upset walked away from the cameraman who was critical of the show.
- (d) **Representationally Rich_Unambiguous**: The actor who was visibly upset walked away from the actress who was critical of the show.

(5) Sentence 2

After a while, he realized it was getting late and took a taxi home.

Unambiguous pronouns have a unique referent and are therefore much less likely to be left unresolved, and even if they are, we must observe the consequence in the comprehension questions: Accuracy for questions explicitly asking about the referent of unambiguous pronouns should be at chance if the pronoun is not resolved during online reading. However, if the results of Experiment 1 were genuinely caused by representational richness, we should still observe an Nref effect in the Baseline condition relative to the Representationally Rich condition even for unambiguous pronouns.

Norming study

As in Experiment 1, we carried out a norming study to measure the relative activation levels of the referential candidates in the ambiguous conditions. Twenty-four participants from the participant pool of the University of California, Davis participated in the experiment. All were native speakers of American English and reported no language-related disorders and no exposure to any language(s) other than English before the age of five. We created 120 sentence pairs such as (4) and (5). As in Experiment 1, the critical discourse fragments included the first sentence (4), plus all the words including and up to the critical pronoun in the second sentence (5), resulting in discourse fragments such as (6).

(6)
- (a) **Baseline_Ambiguous**: The actor walked away from the cameraman. After a while, he …
- (b) **Representationally Rich_Ambiguous**: The actor who was visibly upset walked away from the cameraman who was critical of the show. After a while, he …

Each experimental item occur in one of four versions; two versions captured the two critical conditions as illustrated in 6a and 6b, and two additional versions were made by swapping the linear positions of the two referential candidates to control for any potential intrinsic bias of the NPs. For example, actors are usually more famous people than cameramen and might therefore be more likely to be interpreted as the referent of the pronoun regardless of their syntactic role/position. To control for this, sentences 6a and 6b, for instance, were changed from *actor-cameraman* to *cameraman-actor*. However, we collapsed over this ordering factor when analyzing the data because it was not of theoretical interest. We also added the 40 fillers from Experiment 1 to each list. As in Experiment 1, each list contained only one version of each experimental item. The procedure and coding were identical to that in Experiment 1. We lost 38 (.01% of the) data points because some participants failed to identify a referent for some of the items.

It is also important to mention that we changed many of the NPs in Experiment 2, and even for the items from Experiment 1 that we reused in Experiment 2, we adjusted the relative clauses so that their semantic content would be appropriate for both masculine and feminine referential candidates in the ERP experiment (see below). Also, because unambiguous pronouns make it obvious which referential candidate is being referred to, we did not include the unambiguous conditions in the norming study of Experiment 2. Moreover, the goal of the norming study was to establish the potential effect of the semantic contents of the added relative clauses on the activation levels of the referential

candidates. Since the relative clauses do not vary across the ambiguous and unambiguous conditions, the results within the ambiguous conditions would inform us about the effect of the modifications within the unambiguous conditions as well.

Results of norming study

Table 3 reports the probability of selecting each of the referential candidates. Consistent with Experiment 1 and with previous research (e.g., Arnold, 2001; Fletcher, 1984; Fukumura & Van Gompel, 2010; Givón, 1983; Gordon et al., 1993; Gundel et al., 1993; Karimi & Ferreira, 2016a), there was a general tendency to interpret the ambiguous pronouns as referring to the syntactic subject (NP1) rather than to the syntactic object (NP2) across both conditions. However, this tendency was stronger compared to Experiment 1 (compare Tables 1 and 3). A repeated-measures ANOVA showed a statistically greater probability of referring to NP1 in the Baseline condition compared to the Representationally Rich condition ($F(1,23) = 22.73$, $p < .001$). As in Experiment 1, to assess the effect of participants, items and their interactions with Representational Richness, we also analyzed the data using a logit mixed-effects model with a full random effects structure (Barr et al., 2013), but the results did not change ($\beta = .47$, $SE = .10$, $Z = 4.57$, $p < .001$).

Table 3
The percentages of choosing each referential candidate as the referent of the pronoun in the ambiguous conditions. Experiment 2.

Condition	NP1 preference	NP2 preference
Baseline_Ambiguous	78.1%	21.9%
Representationally Rich_Ambiguous	69.3%	30.7%

Thus, based on the norming results, the relative clauses change the contextual bias (i.e., the probabilities of the two referential interpretations) between the two conditions. Interestingly, however, this change in results creates an opportunity: It establishes a contrast between the predictions of the referential coherence and the retrieval difficulty accounts. Specifically, because contextual bias is weaker in the Representationally Rich than in the Baseline condition, the pronouns following representationally rich (modified) referential candidates should elicit an Nref relative to those following bare (unmodified) referential candidates (Nieuwland & Van Berkum, 2006). This is because the degree to which both referential candidates are entertained and/or the degree of inference needed to arrive at a referential interpretation should be greater in the Representationally Rich compared to the Baseline condition, making the pronouns less coherent, and therefore more difficult to process. However, if referential processing is modulated by ease of memory retrieval alone (i.e., independent of referential coherence), we should observe a main effect of Representational Richness, with an Nref effect for pronouns following bare compared to those following representationally rich referential candidates, as in Experiment 1.

ERP experiment

Method

Participants. Forty-eight participants from the participant pool of the University of California, Davis took part in the ERP experiment.[6] None had participated in Experiment 1 or in either of the norming studies. All participants were right-handed, native speakers of American English and confirmed that they had no language-related or neurological disorders and no exposure to any language(s) other than English before the age of five. They all signed a consent form before taking part in the experiment.

Stimuli, procedure and EEG recording. The stimuli were identical to those employed in the norming study except that the unambiguous conditions (4b and 4d) were added to the design, and the sentences including the critical pronouns were presented in full (see 4 and 5 for an example, and Appendix B for a sample of the stimuli used in this experiment). Thus, the number of experimental lists increased to 8; 4 lists corresponded to the 4 conditions outlined by 4a-d, and 4 additional lists were created by flipping the syntactic roles/positions of the two NPs (see above). However, as with the norming study, we collapsed over this ordering factor when analyzing the data. We also added 82 fillers to each experimental list of the ERP experiment, making the total number of items per list 202 (120 experimental items plus 82 fillers). The fillers contained two or only one referential candidate but in the form of proper names (rather than determiner-NP combinations) and the following pronouns were always either unambiguous or collective (*they*). The sentences in each list were presented to the participants in a fixed randomized order (the order of items was randomized only once, and all versions of an experimental item appeared in the same position in all lists). We also added 40 comprehension questions to the experimental sentences to assess the participants' final interpretations of the ambiguous pronouns (20 questions) as well as their accuracy in resolving the unambiguous pronouns (20 questions). Twenty-six fillers also had comprehension questions tagged to them to obscure the critical sentences. The procedure and the specifics of EEG recording were identical to those in Experiment 1.

Results. Fig. 2 shows the ERPs for the four conditions time-locked to the onset of the critical pronouns. As can be seen in this figure, an Nref effect seems to emerge at about 400 ms after pronoun onset in the Baseline condition relative to the Representationally Rich condition. The analyses (including time windows and the topographical brain regions) were identical to those in Experiment 1. The results are reported in Table 4.

Consistent with the results of Experiment 1, the main effect of Representational Richness was statistically significant in the frontal and central regions in the 400–600 ms window. In the 800–1200 ms window, this effect was significant in all brain regions. The main Ambiguity effect trended towards statistical significance only in the frontal region within both the 400–1200 ms and 800–1200 ms windows. The interaction between Ambiguity and Representational Richness was not significant in any time window and in any of the brain regions.[7]

Although a differential effect of representational richness on ambiguous and unambiguous pronouns would have resulted in an interaction between the two predictors, we still ran separate analyses within the unambiguous conditions and confirmed that the effect of representational richness applied to unambiguous pronouns too (see Appendix C for the results of these analyses).

We also analyzed the behavioral data from this experiment. Accuracy rates for choosing the correct referential candidate for unambiguous pronouns were 87.7% and 87.9% for the Baseline and the Representationally Rich conditions, respectively. This difference was not statistically reliable ($F(1,47) = .02$, $p = .87$). A logit mixed-effects regression model with a full random effects structure (Barr et al., 2013)

[6] We ran more participants in Experiment 2 compared to Experiment 1 because there were more conditions in Experiment 2.

[7] It is important to mention that we initially had 32 participants for Experiment 2 and the interaction between Representational Richness and Ambiguity was significant in that dataset, with the effect of Representational Richness being statistically reliable only within the unambiguous conditions, and the effect of Ambiguity being significant only within the Representationally Rich conditions. However, based on one of the reviewer's recommendation, we decided to run 16 more participants on Experiment 2, and this interaction disappeared. The main effect of Representational Richness was significant in both datasets.

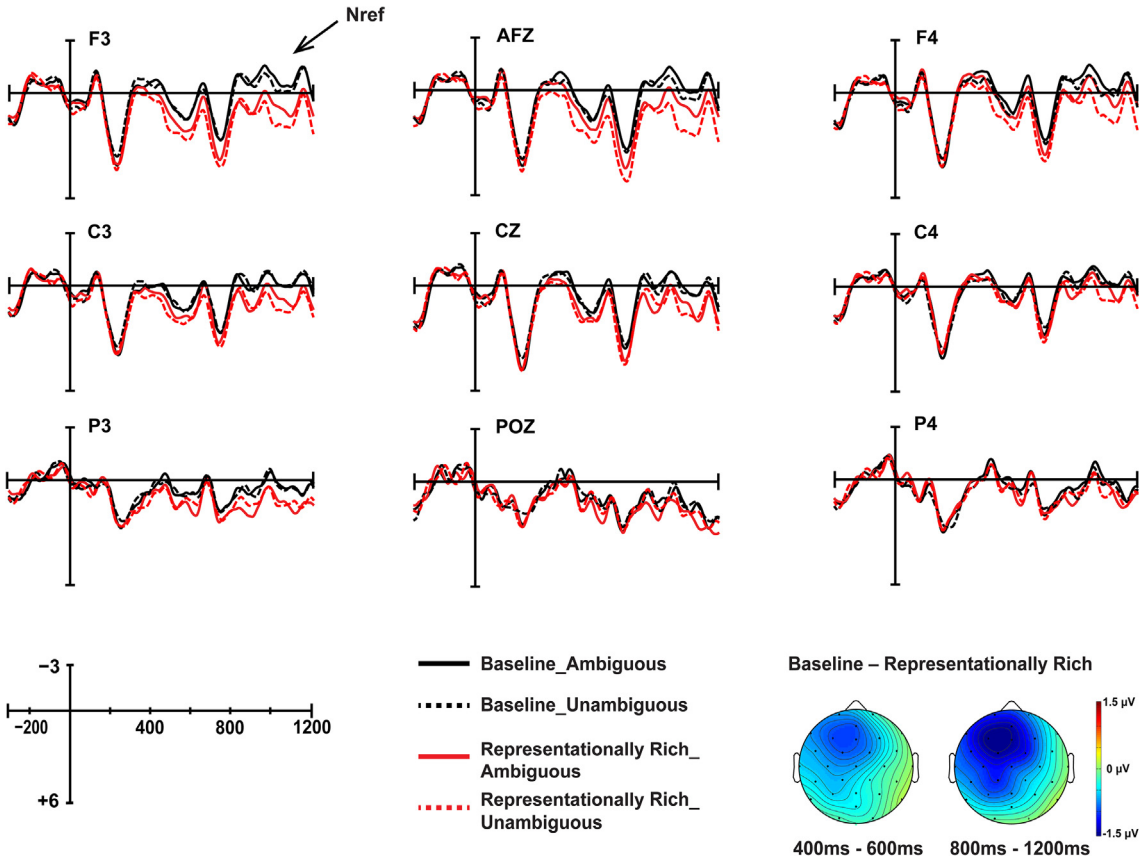

Fig. 2. Grand average brain waveforms time-locked to pronoun onset in the Baseline and Representationally Rich conditions. Experiment 2.

Table 4
ANOVA results in the critical time windows, and within the three brain regions of interest. Experiment 2. "Rep. Richness" denotes "Representational Richness".

Time Window	Predictor	Brain Region	Results
400 ms-600 ms	Ambiguity	Frontal	$F(1,47) = 4.20$, $p = .05$
		Central	$F(1, 47) = 1.58$, $p = .21$
		Posterior	$F(1, 47) = .02$, $p = .88$
	Rep. Richness	Frontal	$F(1,47) = 8.89$, $p = .005$
		Central	$F(1, 47) = 6.38$, $p = .01$
		Posterior	$F(1, 47) = 2.42$, $p = .13$
	Ambiguity × Rep. Richness	Frontal	$F(1,47) = 1.99$, $p = .16$
		Central	$F(1, 47) = .58$, $p = .45$
		Posterior	$F(1, 47) = .03$, $p = .85$
800 ms-1200 ms	Ambiguity	Frontal	$F(1, 47) = 3.56$, $p = .06$
		Central	$F(1, 47) = .85$, $p = .36$
		Posterior	$F(1, 47) = .14$, $p = .70$
	Rep. Richness	Frontal	$F(1,47) = 17.61$, $p < .001$
		Central	$F(1, 47) = 13.51$, $p = .001$
		Posterior	$F(1, 47) = 5.71$, $p = .02$
	Ambiguity × Rep. Richness	Frontal	$F(1,47) = 1.74$, $p = .19$
		Central	$F(1, 47) = 1.43$, $p = .24$
		Posterior	$F(1, 47) = .32$, $p = .57$

further confirmed the ANOVA results ($\beta = .13$, $SE = .28$, $Z = .46$, $p = .64$).

For ambiguous pronouns, the preferences to choose NP1 (i.e., the structurally more prominent referential candidate) as the referent of the pronoun were 72.2% and 70.4% for the Baseline and the Representationally Rich conditions, respectively, which was not a statistically significant difference ($F(1,47) = .17$, $p = .67$). Again, a logit mixed-effects model with a full random-effects structure confirmed these results ($\beta = -.26$, $SE = .28$, $Z = -.91$, $p = .36$).

Discussion. Consistent with the results of Experiment 1, pronouns following unmodified referential candidates elicited an Nref effect relative to pronouns following representationally rich (modified) referential candidates. Crucially, the effect of representational richness, and by extension ease of retrieval, also applied to the unambiguous pronouns, which further suggests that retrieval difficulty can affect referential processing independently from referential coherence. A potential concern with regards to the results of the first experiment was that participants could have strategically left the pronouns unresolved following representationally richer referential

candidates to relieve the burden imposed on computational resources by the semantic and syntactic complexity of modified referential candidates. Since unambiguous pronouns have a unique referent, underspecification is unlikely for them. Moreover, if the unambiguous pronouns were left unresolved, the accuracy for the end-of-trial comprehension questions should have been at chance. Nevertheless, the accuracy rate within the Representationally Rich conditions was 88% (rounded up), suggesting that participants were mostly making the correct referential commitments for unambiguous pronouns in this condition. Experiment 2 also replicated previous studies reporting an Nref effect for ambiguous pronouns relative to unambiguous pronouns (e.g., Nieuwland & Van Berkum, 2006; Van Berkum et al., 1999; see also Nieuwland, et al., 2007; Van Berkum et al., 2003).[8]

Based on Fig. 2, a potential concern about the results of Experiment 2 might be that instead of a larger negativity in the Baseline condition, the results might actually be caused by a larger positivity in the Representationally Rich condition. However, given that this proposed positivity has the same exact morphological distribution as the negativity, and that referential processing has repeatedly been shown to result in a sustained *negativity* (see above), a positivity account is not applicable to our results.

Interestingly, the norming results and the end-of-trial comprehension question results for this experiment were not fully consistent with each other: Whereas the norming results showed a reliably greater NP1 preference for ambiguous pronouns in the Baseline condition, the comprehension question data showed no difference between NP1 and NP2 preferences across the two ambiguous conditions. We can speculate about a few reasons for this discrepancy. First, the two tasks were different: The norming task involved both language comprehension and language production, while the end-of-trial questions involved comprehension only. Second, for the comprehension questions, the participants were presented with the full sentence containing the ambiguous pronoun, whereas these sentences were truncated in the norming task (see above). Thus, it is possible that the information following the pronoun affected how the pronoun was interpreted. However, even with these differences, the results of the comprehension questions show a weak pattern in the same direction as in the norming study. Given the fact that the question-answering task had much less power than did the norming task (48 subjects × 20 items[9] = 960 observations vs. 24 subjects × 120 items = 2880 observations, respectively), the discrepancy could also have simply been due to power differences. In any case, neither task showed stronger contextual bias for the Representationally Rich condition, and therefore the offline findings are not problematic for our ERP results.

General discussion

In two experiments, we observed that pronouns following bare (unmodified) referential candidates (such as *the actor*) elicit a sustained, frontal negativity (i.e., the Nref) relative to pronouns following representationally rich (modified) referential candidates (such as *the actor who was visibly upset*), regardless of whether the pronouns were ambiguous (Experiments 1 and 2) or unambiguous (Experiment 2). Since representationally richer NPs have been shown to be easier to retrieve from memory, and because pronominal ambiguity (i.e., the number of potential referents) and contextual bias (i.e., the probability of potential referential interpretations) were controlled for, our results tease apart referential coherence from retrieval difficulty (i.e., re-accessing the relevant representations), and show that retrieval difficulty modulates the Nref effect independently of referential coherence.

Crucially, and as discussed in the Introduction, a difference in *relative* activation levels of potential referents directly influences referential processing through determining "what" is retrieved (in this case, "which" referential candidate). Specifically, if one referential candidate is more activated than the other, the more activated candidate will be retrieved faster and interpreted as the pronoun's referent, which is consistent with models of retrieval where what is retrieved and the retrieval process itself take place in single step (see McElree, 2006; McElree et al., 2003). However, because the difference in relative activation levels either remained the same (Experiment 1), or was greater in the Baseline than in the Representationally Rich condition (and therefore predicted the opposite of what we found, Experiment 2), our results can only be interpreted in terms of *how* (i.e., the ease with which) the referential candidates were retrieved, and not in terms of *which* candidate was retrieved.

By controlling for discourse factors, our results highlight the role of memory operations during long-distance linguistic dependency resolutions in general and during referential processing in particular, and lend support to cue-based retrieval theories of language processing. According to these theories, linguistic dependencies, including referential dependencies, are resolved when a retrieval cue such as the gender of a pronoun activates the memory items matching that cue (e.g., Lewis & Vasishth, 2005; Lewis et al., 2006). In our current experiments, the gender of the pronoun matched either one or two referential candidates. However, regardless of the number of matching potential referents and the probability of different referential interpretations, pronouns following representationally richer referential candidates were processed more easily relative to pronouns following representationally less rich referential candidates, suggesting that ease of retrieval from memory directly influences how easily a referential dependency is processed.

The role of memory operations during referential processing has also been experimentally confirmed in recent EEG and functional neuroimaging (fMRI) studies. Specifically, Nieuwland and Martin (2017) reported increased power for gamma-band oscillations source localized to the left parietal cortex when a referring expression had a unique preceding referent (e.g., an unambiguous pronoun) compared to when there was no unique referent (an ambiguous pronoun). Since increased power in the gamma-band oscillations in the parietal cortex has been shown to reflect recognition memory (Burke et al., 2014; Gonzalez et al., 2015; Herrmann, Munk, & Engel, 2004; Jacobs, Hwang, Curran, & Kahana, 2006; Jensen, Kaiser, & Lachaux, 2007; Mormann et al., 2005; Osipova et al., 2006), the results were interpreted as reflecting memory recognition processes during referential processing.

Similarly, using functional neuroimaging, Nieuwland et al. (2007) found reliable increases in brain activity in the medial temporal lobe (including the hippocampus) for pronouns that had a unique preceding referent relative to pronouns with multiple potential referents. Because the medial temporal regions have been shown to participate in memory retrieval (Gonzalez et al., 2015; Shannon & Buckner, 2004; also see Wagner, Shannon, Kahn, & Buckner, 2005), the results could be taken as implicating memory retrieval operations during referential processing. Consistent with the above-mentioned theories and experimental findings, our results also highlight the role of memory operations during language comprehension in general and during referential processing in particular.

Our results are also consistent with previous findings showing that representational richness associated with a linguistic item facilitates its subsequent retrieval (Hofmeister, 2011; Hofmeister & Vasishth, 2014; Karimi & Ferreira, 2016a; Karimi et al., 2014; Troyer et al., 2016), and that representational richness overrides the effect of linear distance (and by extension memory decay, Karimi & Ferreira, 2016a; Karimi et al., 2014). How does representational richness aid subsequent retrieval? According to cue-based models of language processing, there are at least two possible mechanisms for the representational richness

[8] Although the ambiguity effect was only marginally significant within the two critical windows separately, it was statistically reliable in a longer window spanning from 400 ms to 1200 ms after pronoun onset in the frontal region (p = .03).

[9] Recall that out of the 40 total comprehension questions, 20 assessed the interpretation of ambiguous pronouns and the 20 assessed the accuracy of interpreting the unambiguous pronouns.

effect. First, it is possible that the extra semantic information predicated of an NP boosts its baseline activation level, thereby facilitating retrieval. Second, more semantic information necessarily renders the associated NP more distinct from other contents of working memory, thereby reducing interference during retrieval (see Gallo, Meadow, Johnson, & Foster, 2008; Jacoby & Craik, 1979; Nairne, 2006; Hofmeister,& Vasishth, 2014). Our current design cannot distinguish between these two possible mechanisms, but it is quite possible that representational richness increases baseline activation levels and reduces retrieval interference simultaneously. Baseline activation of modified NPs could increase because the head noun might be reactivated when the additional semantic information is being processed, and retrieval interference could be reduced because the extra information likely further differentiates the two referential candidates from one another (e.g., Hofmeister, 2011; Lewis & Vasishth, 2005; Lewis et al., 2006), which is the primary function of restrictive relative clauses.

The distinction between retrieval difficulty and referential coherence is relatable to the "bonding" and "resolution" theory of referential processing (Garrod, 1994; Garrod & Sanford, 1994; Garrod & Terras, 2000; Sanford, Garrod, Lucas, & Henderson, 1983; Sturt, 2003), according to which establishing a referential dependency proceeds through two distinct stages: a low-level, automatic bonding stage where a loose and superficial attachment is formed between a pronoun and the representations of potential referents, and a higher-level resolution stage where the possible referential interpretations are evaluated against context, a referential interpretation is adopted, and that interpretation is then semantically integrated into the overall discourse representation. Specifically, because retrieval difficulty concerns the activation and retrieval of referential candidates, it could correspond to the "bonding" stage where all potential referents are initially activated. However, since referential coherence pertains to committing to a referential interpretation that is most consistent with the preceding discourse, it could correspond to the "resolution" stage where a referential interpretation is adopted and integrated with the preceding context.

Importantly, note that although this argument implies that the effect of retrieval difficulty should emerge early and that the referential coherence effect should emerge later, it is incorrect to expect the representational richness effect to arise before the ambiguity effect in our second experiment. This is because, as we mentioned in the Introduction, the ambiguity effect likely results from both retrieval difficulty (due to greater interference during retrieval) and referential coherence (due to entertaining multiple referential interpretations and/or greater inference required to resolve the ambiguity). Consequently, pronominal ambiguity is always bound to change retrieval difficulty (but note that the reverse is not true: retrieval difficulty can vary while ambiguity is controlled for, as shown in this study). To the extent that retrieval difficulty could be equated with the "bonding" stage of referential processing, our results have an important implication for this theory: Referential processing difficulty (at least as reflected by the Nref) could vary as a function of "bonding" difficulty alone, independent of "resolution" difficulty. However, although the distinction between retrieval difficulty and referential coherence might map on the bonding and resolution account of referential processing, our actual ERP data does not and cannot support a distinction between "bonding" and "resolution" stages during referential processing. In fact, it may very well be the case that referential processing involves only a single retrieval process (e.g., see McElree, 2006; McElree et al., 2003). Under such a scenario, our results can be straightforwardly interpreted as showing that representational richness, and by extension retrieval difficulty, facilitates referential processing independent of referential coherence.

Our results also make an important contribution to our understanding of what the Nref indexes. As mentioned above, the results of most past studies are ambiguous with regards to retrieval difficulty vs. referential coherence, and, as a result, the Nref effect reported by those studies could be interpreted in terms of either or both factors. The current results clearly show that the Nref is sensitive to retrieval difficulty independent of referential coherence, suggesting that this ERP effect might be more sensitive to memory operations than to referential coherence of referring expressions *per se*. Thus, it could be the case that referential coherence affects the Nref through inducing variations in retrieval difficulty; low referential coherence complicates, and high referential coherence facilitates the retrieval of relevant representations. In fact, we are not the first to show that the Nref might index memory operations. Barkley, Kluender, and Kutas (2015) showed that frontal negativities are elicited by long-distance linguistic dependencies (including referential as well as syntactic dependencies). Based on their results, the authors argued that sustained, frontal negatives might index a general "association" mechanism whereby a linguistic element is associated with some preceding information to resolve a syntactic or a referential dependency (also see King & Kutas, 1995; Kluender & Kutas, 1993a, 1993b). It is important to note, however, that our claim is not that referential processing difficulty can be entirely reduced to retrieval difficulty; both referential coherence and retrieval difficulty could influence referential processing.

Two aspects of our results merit further discussion. First, the norming results from Experiment 1 were not fully consistent with those from Experiment 2. In Experiment 1, we observed no difference between the two conditions in terms of contextual bias (and therefore the relative activation levels of the two referential candidates), but in Experiment 2, contextual bias was stronger in the Baseline than in the Representationally Rich condition. We argue that this discrepancy is likely due to the fact that there were 40 new items in the norming study of Experiment 2, which means there were also 80 new NPs, 80 new relative clauses and 40 new discourse connecting words. However, the difference in the stimuli was even more substantial because we also changed some of the NPs and some of the relative clauses for the items from Experiment 1 that we re-used in Experiment 2 to make them semantically appropriate given the genders of the potential referential candidates. Overall, 151 NPs were not shared between the two norming studies. It is important to note that even with these changes, the trend in Experiment 1 was in the same direction as that in Experiment 2. Thus, it could be that we simply did not have enough power to reach statistical significance in the norming study of Experiment 1: We had only 80 sentences in Experiment 1 and so 40 items per condition, but 120 sentences in Experiment 2 and therefore 60 items per condition (recall that we normed only the ambiguous conditions in Experiment 2). However, in any case, this discrepancy has no implication for our results or interpretations because each set of norming stimuli was compared to its respective ERP experiment. Second, the ambiguity effect in our second experiment was relatively smaller compared to those reported in previous studies (e.g., Nieuwland & Van Berkum, 2006; Nieuwland et al., 2007; Van Berkum et al., 1999, 2003). We think there are two main reasons for this. First, our items in Experiment 2 exhibited a strong overall NP1 bias and it has been shown that strongly biased items elicit a considerably smaller Nref effect (Nieuwland & Van Berkum, 2006). Second, due to the shortage of definitionally gendered NPs in English (i.e., terms such as *actress*), many of the NPs that we used in our sentences were only stereotypically gendered, weakening our ambiguity manipulation. For instance, we used noun phrases such as *weight lifter* or *hairstylist* for "masculine" and "feminine" referential candidates, respectively (see Appendix B), and although these terms are gender-biased, they are certainly not unambiguous.

Conclusion

We demonstrated that pronouns following bare referential candidates elicit a sustained frontal negativity (the Nref) relative to pronouns following referential candidates that are semantically richer. Crucially, this effect was obtained when the potential factors that could influence referential coherence—that is, ambiguity, contextual bias and feature

match between the pronoun and the preceding potential referents—were held constant. Thus, our results suggest that retrieval difficulty can influence referential processing independently of referential coherence, and provide further evidence for cue-based theories of language processing.

Appendix A. Sample experimental stimuli for Experiment 1. The relative clauses produced the Representationally Rich referential candidates when attached and the Baseline condition when not attached

Item	Sentence 1	Sentence 2
1	The nurse (who had time off work) went to a karaoke bar with the secretary (who was suffering from depression).	In fact, she really liked to sing but did not have a good voice.
2	The monk (who was getting worried about the lack of religious education) argued with the priest (who was thinking of opening a religious school).	It was clear that he did not have any logical case to make.
3	The fisherman (who supported foreign interference in the country) quarreled with the sailor (who opposed stricter water laws).	Truthfully, he never thought the quarrel would eventually end in a fist fight.
4	The colonel (who had failed to prevent the war) betrayed the prince (who was incompetent at running the country).	Granted, he was considering suicide but changed his mind at the last moment.
5	The maid (who was jogging in the park) smiled at the schoolgirl (who was playing on the playground).	No doubt, she was enjoying the nice weather.
6	The witch (who had a magic sword) fought with the heroine (who was suffering from several deep wounds).	Being experienced, she knew very well that speed was of utmost importance.
7	The mechanic (who was working on a new invention) consulted with the engineer (who had picked up a promising project).	According to everyone, he was a smart and hardworking man.
8	The wrestler (who had spent the last six months practicing) challenged the fighter (who had been undefeated for a year).	Immediately, he got changed and started warming up.
9	The surgeon (who had a difficult operation planned) had a question for the doctor (who specialized in head injuries).	Not surprisingly, he had published in many prestigious journals.
10	The anchorwoman (who was persistent to reveal the dirty tricks of the government) telephoned the congresswoman (who was proposing a new anti-corruption law).	Apparently, she has been threatened several times in the past week.
11	The uncle (who was a university lecturer) encouraged the schoolboy (who was going to take part in a science fair).	Later that day, he felt very positive about himself.
12	The granddaughter (who had a sweet tooth) loved to bake cookies with the grandmother (who used a secret family recipe).	Of course, she licked the dough off the spoon before washing it.
13	The girl (who aspired to be on the big screen) gave flowers to the actress (who had won an Oscar).	Apparently, she loved bouquets of orchids and pink roses.
14	The first lady (who was unhappy with the economy) visited the chairwoman (who was trying to solve the tax problem).	Clearly, she was not fully prepared for the meeting.
15	The prisoner (who was on death row) killed the guard (who was taking a nap).	Apparently, he had a fiancée waiting for him back home.
16	The plumber who was struggling to earn a living had a drink with the electrician who received a large bank loan.	Sometimes, he loved sitting on the porch and watching the ocean waves.
17	The executive who liked to think ahead discussed a plan with the mason who was confused about the schedule.	Being very organized, he wrote down the appointments on a calendar.
18	The wizard who was frustrated by the irreparable situation disagreed with the knight who was carefree and lazy.	Suddenly, he came up with a good idea to solve the problem.
19	The spy who had formerly served in the American army greeted the sniper who had been working alone on the mission.	Clearly, he was famous among all his friends for being intelligent.
20	The dancer who was originally from Russia dined with Ms. America who was practicing to make it to the finals.	Everyone agreed that she had a sharp sense of humor.

Appendix B. Sample experimental stimuli for Experiment 2. The relative clauses produced the Representationally Rich referential candidates when attached and the Baseline condition when not attached. The ambiguous and the unambiguous conditions were created by using the NPs preceding and following the slash (/), respectively.

Item	Sentence 1	Sentence 1
1	The nurse (who had time off work) went to a karaoke bar with the secretary/surgeon (who was suffering from depression).	In fact, she really liked to sing but did not have a good voice.
2	The monk (who was worried about the decline of religion in society) argued with the priest/priestess (who lived a sheltered life).	It was clear that he did not have a convincing argument.
3	The fisherman (who sold all of his fish) quarreled with the salesman/saleswoman (who wanted to buy salmon).	Truthfully, he never thought the quarrel would end in a fist fight.
4	The colonel (who had failed to prevent the war) betrayed the prince/princess (who was incompetent at running the country).	Granted, he was considering suicide but changed his mind at the last moment.

5	The maid (who was jogging in the park) smiled at the schoolgirl/altar boy (who was playing on the playground).	No doubt, she was enjoying the fresh air and summer weather.
6	The witch (who had a magic sword) fought with the heroine/hero (who was suffering from several deep wounds).	Being experienced, she knew that it was important to know the opponent's weakness.
7	The doctor (who was planning a difficult operation) consulted with the heart surgeon/nurse practitioner (who specialized in babies).	Apparently, he had no experience administering anesthesia.
8	The engineer (who was working on a new invention) consulted with the businessman/businesswoman (who had money and experience).	According to everyone, he was a little bit unconventional and didn't look the part.
9	The wrestler (who was a three-time champion) was training with the fighter/female fighter (who was undefeated for the past two years).	After the workout, he went to the locker room and got changed.
10	The bishop (who opposed the government) had a secret meeting with the congressman/congresswoman (who was proposing a new anti-corruption law).	Apparently, he had been receiving several threating calls in the past week.
11	The rabbi (who taught Hebrew at the temple) had a short conversation with the schoolboy/schoolgirl (who was going to take part in the science fair).	At 3o'clock, he had a late lunch of fish and chips.
12	The excited girl (who aspired to be on the big screen) embraced the famous actress/actor (who had won an Oscar).	Suddenly, she was distracted by the fire engine that had just come around the corner.
13	The little girl (who had a sweet tooth) loved to bake cookies with the grandmother/grandfather (who knew a lot of recipes).	Of course, she always tasted the batter even though it had raw eggs.
14	The prisoner (who was on death row) was wrestling with the guard/policewoman (who was on the night shift).	While on the ground, he felt around for something to use as a weapon.
15	The foreman (who was unhappy with the minimum wage) visited the governor/governess (who was trying to solve the budget problem).	Clearly, he was not fully prepared for the meeting.
16	The plumber (who had signed a good deal) had a drink with the electrician/kitchen maid (who received a large bank loan).	Sometimes, he would drink a little too much while celebrating.
17	The executive (who liked to think ahead) discussed the plan with the chairman/chairwoman (who was interested in the outcome of the project).	During the meeting, he took a lot of handwritten notes.
18	The knight (who was frustrated by the situation) disagreed with the wizard/witch (who was carefree and lazy).	Suddenly, he came up with a good idea to solve the problem.
19	The spy (who had formerly served in the American army) greeted the sniper/tall woman (who was smoking in the corner).	Not surprisingly, he was known for being smart and ruthless.
20	The supermodel (who was originally from Russia) dined with the pageant girl/bodybuilder (who was practicing to make it to the finals).	Everyone agreed that she was a natural beauty even without any makeup.
21	The carpenter (who was very experienced) worked with the welder/interior designer (who was good at his job).	There was a chance that he would try to overcharge the business owner.
22	The lion tamer (who worked hard yesterday) was resting with the clown/belly dancer (who just finished his performance).	Tomorrow, he needed to wake up early to practice his routine.
23	The ninja (who had been practicing for months) challenged the Kung Fu master/female black belt (who knew advanced techniques).	Unfortunately, he had no idea that the fight was being filmed.
24	As it turned out, the pretty teacher (who was new to the school) had an affair with the beautiful student/muscular student (who was struggling with math).	When the principal found out, she was able to deny the entire thing.
25	The lumberjack (who was short but sturdy) made friends with the undertaker/virgin (who had a nice smile).	According to our neighbors, he was a nice person but a little odd.
26	The thief (who had botched the robbery) was grappling with the man/woman (who was awakened by the noise).	It was clear that he knew some martial arts, given his swift moves.
27	The rancher (who was wearing thick boots) walked along a cliff with the country boy/country girl (who knew the area).	As you might expect, he didn't enjoy the heavy wind and rain.
28	The holy man (who was holding a sharp stick) confronted the vampire/vampiress (who stepped out of the coffin).	You could tell that he was enraged by the look in his eyes.
29	The CEO (who was excited about the opportunity) negotiated with the Englishman/Englishwoman (who was selling his business).	It was obvious that he got the better end of the deal.
30	The sound engineer (who was very meticulous) tested the streaming audio with the anchorman/anchorwoman (who was excited about the interview).	The night before, he had a bad dream that something would go terribly wrong.

Appendix C. ANOVA results for the representational richness effect within the unambiguous conditions. Experiment 2

Time Window	Brain Area	Results
400 ms-600 ms	Frontal	$F(1,47) = 8.72$, $p = .005$
	Central	$F(1,47) = 5.21$, $p = .03$
	Posterior	$F(1,47) = .63$, $p = .43$
800 ms-1200 ms	Frontal	$F(1,47) = 15.39$, $p < .001$
	Central	$F(1,47) = 12.84$, $p = .001$
	Posterior	$F(1,47) = 1.69$, $p = .20$

References

Ariel, M. (1990). *Accessing noun-phrase antecedents.* London, UK: Routledge.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes, 31*, 137–162.

Barkley, C., Kluender, R., & Kutas, M. (2015). Referential processing in the human brain: An Event-Related Potential (ERP) study. *Brain research, 1629*, 143–159.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255–278.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*, 47–67.

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua, 118*, 172–190.

Burke, J. F., Sharan, A. D., Sperling, M. R., Ramayya, A. G., Evans, J. J., Healey, M. K., ... Kahana, M. J. (2014). Theta and high-frequency activity mark spontaneous recall of episodic memories. *Journal of Neuroscience, 34*, 11355–11365.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research, 1146*, 172–184.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *The Quarterly Journal of Experimental Psychology, 69*, 817–828.

Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior, 22*, 121–132.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass, 1*, 71–83.

Fiorentino, R., Covey, L., & Gabriele, A. (2018). Individual differences in the processing of referential dependencies: Evidence from event-related potentials. *Neuroscience Letters, 673*, 79–84.

Fletcher, C. R. (1984). Markedness and topic continuity in discourse processing. *Journal of Verbal Learning and Verbal Behavior, 23*, 487–493.

Fukumura, K., & Van Gompel, R. P. G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language, 62*, 52–66.

Fukumura, K., & Van Gompel, R. P. G. (2011). The effects of animacy in the choice of referring expressions. *Language and Cognitive Processes, 26*, 1472–1504.

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58*, 1095–1111.

Garrod, S. (1994). Resolving pronouns and other anaphoric devices. In C. Clifton, L. Frazier, & K. Rayner (Eds.). *Perspectives on sentence processing* (pp. 339–359). Englewood, NJ: Erlbaum.

Garrod, S., & Sanford, A. J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In M. A. Gernsbacher (Ed.). *Handbook of psycholinguistics* (pp. 675–698). San Diego, CA: Academic Press.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language, 42*, 526–544.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition, 32*, 99–156.

Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes, 26*, 67–86.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. P. Marantz, & W. O'Neil (Eds.). *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.). *Topic continuity in discourse: A quantitative cross language study* (pp. 1–41). Amsterdam, The Netherlands: Benjamins.

Gonzalez, A., Hutchinson, J. B., Uncapher, M. R., Chen, J., LaRocque, K. F., Foster, B. L., ... Wagner, A. D. (2015). Electrocorticography reveals the temporal dynamics of posterior parietal cortical activity during recognition memory decisions. *Proceedings of the National Academy of Sciences, 112*, 11066–11071.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17*, 311–347.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of anaphoric expressions in discourse. *Language, 69*, 274–307.

Herrmann, C. S., Munk, M. H., & Engel, A. K. (2004). Cognitive functions of gamma-band activity: Memory match and utilization. *Trends in Cognitive Sciences, 8*, 347–355.

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes, 26*, 376–405.

Hofmeister, P., & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology, 5*, 1237. http://dx.doi.org/10.3389/fpsyg.2014.01237.

Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making. *Neuroimage, 32*, 978–987.

Jacoby, L. L., & Craik, F. I. (1979). Effects of elaboration of processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak, & F. I. M. Craik (Eds.). *Levels of processing in human memory* (pp. 1–22). Hillsdale, NJ, USA: Erlbaum.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language, 94*, 316–339.

Jensen, O., Kaiser, J., & Lachaux, J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences, 30*, 317–324.

Johns, C. L., Gordon, P. C., Long, D. L., & Swaab, T. Y. (2014). Memory availability and referential access. *Language, Cognition and Neuroscience, 29*, 60–87.

Karimi, H., & Ferreira, F. (2016a). Informativity renders a referent more accessible: Evidence from eyetracking. *Psychonomic Bulletin & Review, 23*, 507–525.

Karimi, H., & Ferreira, F. (2016b). Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology, 69*, 1013–1040.

Karimi, H., Fukumura, K., Ferreira, F., & Pickering, M. J. (2014). The effect of noun phrase length on the form of referring expressions. *Memory and Cognition, 42*, 993–1009.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25*(1), 1–44.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*, 376–395.

Kluender, R., & Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*, 196–214.

Kluender, R., & Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes, 8*, 573–633.

Koornneef, A. W., & Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from selfpaced reading and eye tracking. *Journal of Memory and Language, 54*, 445–465.

Lamers, M. J., Jansma, B. M., Hammer, A., & Münte, T. F. (2006). Neural correlates of semantic and syntactic processes in the comprehension of case marked pronouns: Evidence from German and Dutch. *BMC Neuroscience, 7*, 23.

Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical repetition: Mechanisms of discourse integration. *Memory & Cognition, 35*, 801–815.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*, 447–454.

Lucas, M. M., Tanenhaus, M. K., & Carlson, G. N. (1990). Levels of representation in the interpretation of anaphoric reference and instrument inference. *Memory & Cognition, 18*, 611–631.

MacDonald, M. C., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language, 29*, 469–492.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1231–1239.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage, 59*, 1859–1869.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain and Language, 135*, 42–51.

McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation, 46*, 155–200.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*, 67–91.

Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., & Elger, C. E. (2005). Phase/amplitude reset and theta–gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. *Hippocampus, 15*, 890–900.

Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt, & J. B. Worthen (Eds.). *Distinctiveness and memory* (pp. 27–46). Oxford, UK: Oxford University Press.

Nieuwland, M. S. (2014). "Who's he?" Event-related brain potentials and unbound pronouns. *Journal of Memory and Language, 76*, 1–28.

Nieuwland, M. S., & Martin, A. E. (2017). Neural oscillations and a nascent corticohippocampal theory of reference. *Journal of Cognitive Neuroscience, 29*, 896–910.

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience, 19*, 228–236.

Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. *NeuroImage, 37*, 993–1004.

Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research, 1118*, 155–167.

Nieuwland, M. S., & Van Berkum, J. J. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language, 106*, 119–131.

Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience, 26*, 7523–7531.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language, 42*, 168–182.

Qiu, L., Swaab, T. Y., Chen, H. C., & Wang, S. (2012). The role of gender information in pronoun resolution: Evidence from Chinese. *PLoS ONE, 7*, e36156.

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience, 29*, 912–927.
Rosenbach, A. (2008). Animacy and grammatical variation—Findings from English genitive variation. *Lingua, 118*, 151–172.
Sanford, A. J., & Garrod, S. (1989). What, when and how: Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes, 4*, 235–263.
Sanford, A. J., & Garrod, S. C. (2005). Memory-based approaches and beyond. *Discourse Processes, 39*, 205–224.
Sanford, A., Garrod, S., Lucas, A., & Henderson, R. (1983). Pronouns without explicit antecedents. *Journal of Semantics, 2*, 303–318.
Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences, 6*, 382–386.
Shannon, B. J., & Buckner, R. L. (2004). Functional-anatomic correlates of memory retrieval that suggest nontraditional processing roles for multiple distinct regions within posterior parietal cortex. *Journal of Neuroscience, 24*, 10084–10092.
Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*, 542–562.
Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience, 16*, 715–726.
Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition, 36*(1), 201–216.
Troyer, M., Hofmeister, P., & Kutas, M. (2016). Elaboration over a discourse facilitates retrieval in sentence processing. *Frontiers in Psychology, 7*. http://dx.doi.org/10.3389/fpsyg.2016.00374.
Van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147–182.
Van Berkum, J. J., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*, 235–248.
Van Berkum, J. J., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158–171.
Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55*, 157–166.
Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences, 9*, 445–453.
Yamashita, H., & Chang, F. (2001). "Long before short" preference in the production of a head-final language. *Cognition, 81*, B45–B55.